PB85-241461

Choosing Cost-Effective QA/QC (Quality Assurance/Quality Control) Programs for Chemical Analysis

Radian Corp., Austin, TX

Prepared for

Environmental Monitoring and Support
Lab.-Cincinnati, OH

Aug 85

U.S. Department of Commerce
National Technical Information Service
NTIS

PB85-241461

EPA/600/4-85/056
August 1985

CHOOSING COST-EFFECTIVE QA/QC PROGRAMS FOR CHEMICAL ANALYSIS

Lloyd P. Provost
Robert S. Elder
Radian Corporation
8501 Mo-Pac Blvd.
Austin, Texas 78766

Final Report
EPA Contract No. CI-68-03-2995

James E. Longbottom/Stephen Billets, Project Officers
Environmental Monitoring and Support Laboratory
U.S. Environmental Protection Agency
Cincinnati, Ohio

Environmental Monitoring and Support Laboratory
Office of Research and Development
U.S. Environmental Protection Agency
Cincinnati, OH 45268

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA/600/4-85/056	2.	3. RECIPIENT'S ACCESSION NO. PB85 241461
4. TITLE AND SUBTITLE Choosing Cost-Effective QA/QC Programs for Chemical Analysis		5. REPORT DATE August 1985
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Lloyd P. Provost and Robert S. Elder		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Radian Corporation 8501 Mo-Pac Blvd. Austin, TX 78766		10. PROGRAM ELEMENT NO. CBE1BC
		11. CONTRACT/GRANT NO. 68-03-2995
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Monitoring and Support Laboratory 26 W. St. Clair St. Cincinnati, Ohio 45268		13. TYPE OF REPORT AND PERIOD COVERED Final May 1980-Jan 1985
		14. SPONSORING AGENCY CODE EPA 600/06

15. SUPPLEMENTARY NOTES

16. ABSTRACT

The primary objective of this report is to provide guidance for choosing cost-effective QA/QC programs for chemical laboratories. It describes general principles of QA/QC, the specific tools available, and the information needed to choose appropriate tools for specific needs. The report does not give detailed discussions of how to apply each quality control tool; references are given for more detailed information.

The major tasks of this project were:

- A literature search to identify current QA/QC practices for inorganic and organic chemical methods
- An evaluation of ongoing quality assurance programs
- Development of a model to determine the type and level of QA/QC effort required for various uses of particular analytical methods.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT Release to public	19. SECURITY CLASS *(This Report)* Unclassified	21. NO. OF PAGES 153
	20. SECURITY CLASS *(This page)* Unclassified	22. PRICE

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

FOREWORD

Environmental measurements are required to determine the quality of ambient waters and the character of waste effluents. The Environmental Monitoring and Support Laboratory (EMSL)-Cincinnati conducts research to:

- Develop and evaluate techniques to measure the presence and concentration of physical, chemical, and radiological pollutants in water, wastewater, bottom sediments, and solid waste.
- Investigate methods for the concentration, recovery, and identification of viruses, bacteria, and other microorganisms in water.
- Conduct studies to determine the responses of aquatic organisms to water quality.
- Conduct an Agency-wide quality assurance program to assure standardization and quality control of systems for monitoring water and wastewater.

This publication, Choosing Cost-Effective QA/QC Programs for Chemical Analysis, reports the results of EPA's literature search and review, visitations to government and private laboratories, and analysis of QA/QC issues and data. Federal agencies, states, municipalities, universities, private laboratories, and industry should find this study useful in developing QA/QC programs for environmental analysis.

Robert L. Booth, Director

ABSTRACT

This report was submitted in fulfillment of contract number 68-03-2995 by Radian Corporation under the sponsorship of the U.S. Environmental Protection Agency (USEPA). The report covers a period from November, 1980 to January, 1985.

The Environmental Monitoring and Support Laboratory, U.S. Environmental Protection Agency, Cincinnati, has the responsibility for developing quality control procedures which could be incorporated into a laboratory quality assurance program designed to support its ongoing analytical method development research and monitoring programs. Radian Corporation was contracted to review current quality assurance (QA) and quality control (QC) programs and develop guidelines for QA/QC practices for the USEPA 600 series methods for chemical analysis of toxic organic pollutants (proposed December 3, 1979, Federal Register, 44 (233), pp. 69464-69575). Use of the proposed analytical procedures "would be required for filing applications for National Pollutant Discharge Elimination System (NPDES) permits, for State certifications, and for compliance monitoring under the Clean Water Act."

The major tasks of this project were:

- A literature search to identify current QA/QC practices for inorganic and organic chemical methods
- An evaluation of ongoing quality assurance programs
- Development of a model to determine the type and level of QA/QC effort required for various uses of particular analytical methods

The primary objective of this report is to provide guidance for choosing cost-effective QA/QC programs for chemical laboratories. It describes general principles of QA/QC, the specific tools available, and the information needed to choose appropriate tools for specific needs. The report does not give detailed discussions of how to apply each quality control tool; references are given for more detailed information.

The report is not targeted at any particular type of laboratory (e.g., EPA contractor) or any specific analytical method. The USEPA 600 series methods are used for exemplary purposes in the report.

CONTENTS

Foreword.. iii
Abstract.. iv
Figures... vi
Table... vii

1. Introduction................................. 1
   Project Background........................... 1
   Contents of This Report...................... 1
   Bibliography................................. 2
2. General QA/QC Principles..................... 4
   Definition of Quality as Fitness for Use..... 4
   Total Quality Control........................ 5
   Resource Allocation.......................... 6
   Process Control.............................. 8
   Measures of Analytical Quality............... 10
   Simplicity................................... 11
   References................................... 13
3. QA/QC Tools.................................. 15
   Blanks....................................... 18
   Calibration.................................. 20
   Control Charts............................... 26
   Interlaboratory Studies...................... 44
   Material Controls............................ 48
   Method Development........................... 50
   Performance and System Audits................ 55
   Reference Materials.......................... 57
   Replication.................................. 59
   Sampling Procedures.......................... 62
   Spike-Recovery Studies....................... 65
   Study Planning............................... 68
   Surrogate Compounds.......................... 69
   Validation................................... 71
4. Measuring QA/QC Cost Effectiveness........... 73
   Achieving QC Targets......................... 74
   End-Use Quality.............................. 85
   Concluding Remarks on QA/QC Effectiveness.... 107
   References................................... 108
5. Choosing Cost-Effective QA/QC Programs....... 111
   Minimal QA/QC Programs....................... 112
   Additional QA/QC Efforts..................... 119
   References................................... 125

Appendices

A. Skip-Lot Procedures.......................... 127
B. Design and Analysis of Spike-Recovery Studies..... 131

FIGURES

Number		Page
3-1	$\bar{X}$ Control Chart Illustration......................	29
3-2	Multivariate Chart Example........................	41
4-1	Chance of Detecting a Specified Bias in m Points on an $\bar{X}$ Control Chart with 3σ Action Limits.......	76
4-2	Chance of Detecting a Change in Precision in m Points on a Range Control Chart..................	80
4-3	Number of QC Tests Required to Detect a Quality Problem..	82
4-4	Nomograph to Determine the Number of Replications Required to Achieve a Specified Maximum Error.....	88
4-5	Nomograph to Determine the Optimum Number of Replications in the Second Stage of a Two-Stage Procedure..	94
4-6	OC Curve Example.................................	96
4-7	Example of Operating Characteristic (OC) Curves for Method 607..................................	99
4-8	Cost-Effectiveness of Replication................	102
4-9	Cost-Effectiveness of Bias Correction............	103
5-1	Quality Control Organizations for Laboratories....	114
5-2	Steps Involved in Tailoring a Quality Control Program for a Particular Use.....................	121

TABLES

Number		Page
3-1	Numbers of Compounds Covered by USEPA Wastewater Methods	32
3-2	Bonferroni Z-Values for Multiple Tests	34
3-3	Parameters for χ^2 Control Charts	37
3-4	Simulated Percent Recoveries for Example	40
3-5	Factors for Ruggedness Test for Method 625	52
3-6	Design Matrix for Method 625 Ruggedness Test	54
4-1	Formulas for Computing OC Curves	98

SECTION 1

INTRODUCTION

PROJECT BACKGROUND

The Environmental Monitoring and Support Laboratory, U.S. Environmental Protection Agency, Cincinnati office, has the responsibility for developing quality control procedures which could be incorporated into a laboratory quality assurance program designed to support its ongoing analytical method development research and monitoring programs. Radian Corporation was contracted to review current QA/QC programs* and develop guidelines for QA/QC practices for the USEPA 600 series methods for chemical analysis of toxic organic pollutants.

CONTENTS OF THIS REPORT

The primary objective of this report is to provide guidance for choosing cost-effective analytical QA/QC programs. To this end, the report describes:

- General principles of quality control that provide a conceptual framework for QA/QC program design (Section 2)
- Alternate tools available for analytical quality control with qualitative guidance for ensuring their effectiveness (Section 3)

---

*QA = quality assurance, the system of activities whose purpose is to provide assurance that the quality-control job is being done effectively.

QC = quality control, the system of activities whose purpose is to provide a quality of product or service that meets the needs of users.

- Decision-directing formulae for determining the type and frequency of QA/QC activities needed to achieve specified quality targets (Section 4)
- Formulae for determining quality targets appropriate for particular end-use needs (Section 4)
- Procedures for evaluating and improving the cost-effectiveness of quality assurance programs (Section 4)
- An approach to structuring QA/QC programs for methods whose results may be put to different uses (Section 5).

The decision-directing formulae are presented graphically to facilitate use. They are illustrated with examples based on self-monitoring or regulatory applications.

BIBLIOGRAPHY

The general sources listed below were particularly useful in preparing this report. Other sources found useful for specific purposes are cited later in the report.

1. ACS Committee on Environmental Improvement. "Principles of Environmental Analysis." Analytical Chemistry, 55, 1983, pp. 2210-2218.

2. DeVoe, J. R., ed., Validation of the Measurement Process, ACS Symposium Series 63, American Chemical Society, Washington, D.C., 1977.

3. Environmental Monitoring and Support Laboratory, Handbook for Analytical Quality Control in Water and Wastewater Laboratories, EPA-600/4-79-019, U.S. Environmental Protection Agency, Office of Research and Development, Cincinnati, Ohio, 1979.

4. Environmental Monitoring and Support Laboratory, *Quality Assurance Handbook for Air Pollution Measurement Systems, Volume I - Principles*, EPA-600/9-76-005, U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, North Carolina, 1976.

5. Garfield, F. M., N. Palmer and G. Schwartzman, eds., *Optimizing Chemical Laboratory Performance Through the Application of Quality Assurance Principles*, Association of Official Analytical Chemists, Arlington, Virginia, 1980.

6. Juran, J. M. and F. M. Gryna, *Quality Planning and Analysis*, McGraw-Hill, New York, 1970.

7. Ku, H. H., ed., *Precision Measurement and Calibration*, NBS Special Publication 300, U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., 1969.

8. LaFleur, P. D., ed., *Accuracy in Trace Analysis - Volume I*, NBS Special Publication 422, U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., 1976.

9. Liteanu, C. and I. Rica, *Statistical Theory and Methodology of Trace Analysis*. Holsted Press, New York, 1980.

10. MacDougall, D., et al., "Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry," *Analytical Chemistry*, 52(14), 1980, pp.2242-2249.

11. Massart, D. L., A. Dijkstra, and L. Kaufman, *Evaluation and Optimization of Laboratory Methods and Analytical Procedures*, Elsevier Scientific Publishing Company, Amsterdam, The Netherlands, 1978.

12. Wilson, A. L., "Approach for Achieving Comparable Analytical Results From a Number of Laboratories," *The Analyst*, 104 (1237), 1979, pp.273-289.

SECTION 2

GENERAL QA/QC PRINCIPLES

This section discusses general concepts and approaches found useful by the quality control profession in over fifty years of accumulated experience. The principles described are qualitative, but serve as guides to quantitative evaluations of QA/QC programs. They also are helpful guides in starting quality programs for new test methods (where information needed for quantitative decisions may be lacking) or for choosing minimal programs for methods whose results are put to many uses.

DEFINITION OF QUALITY AS FITNESS FOR USE

Juran (1) says that "Of all concepts in the quality function..., none is so far-reaching or vital as 'fitness for use.'" The interpretation of quality as fitness for use rather than conformance to specifications (another common interpretation) shows the importance of:

- Basing quality targets on end-use needs
- Building flexibility into QC programs for products with multiple uses.

Methods for relating quality control objectives to end-use needs are discussed in Section 4; methods for building flexibility into required quality control programs are provided in Section 5.

It is common practice to put the products of analytical chemistry - test results - to many uses. For example, the EPA Handbook for Analytical Quality Control in Water and Wastewater Laboratories (2) lists seven uses of analytical data:

- Planning
- Permitting
- Compliance
- Enforcement
- Design
- Process Control
- Research and Development

These uses involve such activities as characterizing variability in pollutant concentrations, comparing measured concentrations to regulatory limits, comparing concentrations from different treatment systems, and studying changes in pollutant concentrations over time. Because of the many uses that can be made of environmental data, it is not possible to design a single QA/QC program that will be cost-effective for every application. Fitness for use requires that at least some aspects of a quality program be tailored to the circumstances and needs of each problem. For example, low level contamination may not be important when comparing concentrations from different treatment systems, but may be critical when comparing concentrations to regulatory limits.

TOTAL QUALITY CONTROL

QA/QC programs are most effective when they are comprehensive, that is when they are involved in every stage from the development to the use of a product or service (3 - 4). A comprehensive analytical QA/QC program involves method developers, producers of materials and equipment, laboratory managers, analysts, quality control personnel and users of analytical results. The work of these people must be coordinated to ensure that all parties are aware of and carry out their responsibilities and communicate their knowledge to others in the program (4). Juran and Gryna (5) provide an excellent discussion of organization for quality.

Much of the rest of this report is devoted to QA/QC activities and responsibilities in the application of a test method. However, several steps must be taken in the development and implementation of a method if routine quality control activities are to be fruitful. Method development should include ruggedness and interlaboratory testing (both are discussed in Section 3). Method implementation in a laboratory should include materials and equipment testing, analyst training and evaluation, and method validation (these topics also are discussed in Section 3).

Quality-related efforts in method development and implementation are important because they can prevent problems from occurring in method application. In addition, they provide cost and quality information needed to design effective QA/QC programs for particular problems.

RESOURCE ALLOCATION

An efficient QA/QC program allocates resources to problems in proportion to their seriousness. To design an efficient program, therefore, one must determine the relative seriousness of quality-related problems; for example, estimate the proportion of total variation attributable to each component of a measurement system. The relative costs of QA/QC activities also must be identified. Designing an effective program usually is an iterative process, because understanding of costs and sources of problems in a test method increases with experience.

The American Society for Quality Control (ASQC) publishes a useful guide to evaluating quality costs (6). It separates these costs into three categories:

- Prevention - costs of activities to prevent poor quality from occurring (e.g., training costs, costs of ruggedness testing).
- Appraisal - costs of evaluating outgoing quality (e.g., costs of estimating recovery from spiked samples).
- Failure - direct and indirect costs of laboratory errors (e.g., costs of rerunning improperly analyzed samples, costs of incorrect regulatory decisions).

Detailed discussions of the three types of costs are given in references (5 - 7); included are methods of identifying opportunities for savings (either by eliminating activities whose costs are disproportionate to their benefits, or by introducing activities with high cost-effectiveness). Failure costs can be the most serious but also may be most difficult to quantify.

Cost improvements usually come through preventive actions aimed at specific quality problems (6). Thus, it is important to use quality appraisal data not only to document out-going quality, but to identify the existence both of correctable problems whose causes can be eliminated and prevented from recurring, and of exceptional quality whose causes can be identified and disseminated. Preventive measures are cost effective because "doing it right the first time" saves reworking, trouble-shooting and other failure costs.

Procedures have been developed to tie the level of quality control effort automatically to the size of the quality problem. Skip-lot procedures, for example, are designed to reduce QC testing when recent quality has been good, and to increase testing if quality subsequently deteriorates. Such procedures are appealing because they provide an incentive for maintaining good performance. Skip-lot procedures are described in Appendix A.

PROCESS CONTROL

The idea of process control was developed in the manufacturing industries but has been found useful in many other applications. The need for process control arises from the universal fact that quality varies among items produced. Since it generally is impractical to measure the quality of every item, quality usually is described in terms of properties of its statistical distribution (e.g., by the average and standard deviation estimated from a sample from the distribution). An important consideration with this method is that stable quality distributions do not naturally exist in a laboratory, and when there is no stable distribution, simple descriptive methods do not apply.

One purpose of process control is to ensure the existence of a stable quality distribution. This goal is achieved by separating causes of variation into "common causes" (variation associated with the analytical system) and "special causes" and dealing appropriately with each. Common causes, often called chance causes, typically result in relatively small, random errors. (Random changes in the testing environment, test procedure and instrument performance are examples of chance causes.) Little can be done to reduce chance variation outside of making basic changes in the analytical process. However, variation due to common causes follows statistical laws; it is predictable in a statistical sense. In process control, therefore, when quality varies in a manner that might reasonably be produced by chance causes (i.e., conforms to predicted statistical patterns), it is assumed that no special causes are present and the production process is "in control." When quality variation does not conform to predicted statistical patterns, on the other hand, it is concluded that at least one special cause (often called assignable cause) is present. Special causes typically result in relatively large, systematic errors. (Examples of such causes are systematic differences among workers, equipment, or materials, contamination of reagents, and systematic changes in environmental conditions over time.) Detecting, eliminating and preventing the recurrence of special causes are basic process-control activities.* The ASQC cost-reducing guide (6) describes process control as a "vital part of a prevention-oriented quality system."

Wernimont (10) and Mandel (11) give excellent descriptions of the importance of process control in measurement. To appreciate this importance, one must make a distinction between:

- A measurement method - the specifications of the equipment and materials to be used, the operations to be performed and the conditions under which they are to be carried out (sometimes called the "protocol")
- Measurement processes - realizations of the measurement method in different conditions and circumstances (i.e., the application of the method in different laboratories).

Eisenhart (12) stresses the need to view measurement as a production process and the importance of maintaining the process in a state of statistical control. Murphy (13) says:

> Capability of control means that either the measurements are the product of an identifiable statistical universe... or, if not, the physical causes preventing such an identification may themselves be identified and, if desired, isolated and suppressed. Incapability of control implies that the results of measurement are not to be trusted as indications of the physical property at hand - in short, we are not in any verifiable sense measuring anything.

---

*See (8) and (9) for more details on the concept of process control. Control charts are a tool commonly used to describe, attain and maintain process control. Their use is described in Section 3.

This often-quoted statement makes clear the importance of process control in analytical QA/QC.

MEASURES OF ANALYTICAL QUALITY

There are several different measures of analytical quality corresponding to the different kinds of errors that can occur in analytical work:

- Systematic errors
- Random errors
- Detection errors (false positives or false negatives)
- Total failures (instances when no result can be reported because of lost samples or other mistakes).

The importance of each kind of error depends on the application and on the magnitude or frequency of the error. For example, detection errors may be of primary concern in screening studies, whereas systematic errors may be of primary concern in estimating pollutant concentrations.

The following concepts can be used to quantify the seriousness of the errors in a particular measurement process:

- **Bias** - the direction and amount by which measurements tend to differ from the true value of the quantity of interest; a measure of systematic error
- **Precision** - the degree of mutual agreement between independent measurements made under prescribed like conditions; a measure of random error

- Sensitivity - the probability of detecting a compound when it is present in a sample; a measure of infrequency of false negatives
- Specificity - the probability of not detecting a compound when it is not present in a sample; a measure of infrequency of false positives
- Completeness - the amount of valid data obtained from a measurement system compared to the amount expected to be obtained under normal operations; a measure of infrequency of total failures.

An in-control measurement process must exist for these concepts to be uniquely quantifiable; otherwise, estimates obtained at one time will not reflect quality at other times.

There is no consensus on most of the definitions given above. For example, bias, as defined, is sometimes referred to as accuracy,* and there are several definitions of sensitivity (15). Because of the lack of agreement on terminology, published "precision and accuracy" data should be accompanied by a description of how computations were done and the circumstances under which results were obtained (15).

SIMPLICITY

Harold F. Dodge, near the end of a distinguished career in quality control, noted that "...there was one thing that seemed to stand out, and it was this: If you want a method or system used, keep it simple!" (16). Dodge's advice should be kept in mind when adapting QA/QC procedures to analytical quality control. For example, the effectiveness of the simple plotting techniques called control charts can be greatly impaired if charts are required for too many parameters.

---

*ASTM (14) recommends the use of the terms bias and precision to describe the accuracy of a measurement process. In this usage, accuracy reflects both systematic and random errors.

Hamaker (17) stresses the importance of simplicity in discussing mathematical models used to optimize quality programs. He notes that the many publications treating quality control as an economic problem "have never been applied on a scale worth mentioning." He gives the following reasons for this failure:

- The techniques are too complicated for practical purposes. "Simplicity is the keystone of success..."
- They are based on unrealistic assumptions (which often must be made to ensure mathematical tractability).

Hamaker's comments show the futility of attempting to formulate a mathematical model encompassing all aspects of laboratory quality control.

The practical approach to designing cost-effective quality control programs involves:

- Identifying and setting tolerances and controls for critical steps in the test method
- Setting quality targets based on end-use needs
- Using statistical methods and models to choose effective quality control tools to achieve quality targets
- Periodically reviewing the effectiveness of the QA/QC system and making improvements based on experience.

Potential quality control tools are discussed in the next section, and methods for choosing cost-effective quality control programs are discussed in Sections 4 and 5. The time-tested principles discussed in this section provide a conceptual framework for these later sections of the report.

REFERENCES

1. Juran, J. M., (ed.); Quality Control Handbook, 3rd edition, McGraw-Hill, New York, 1974, p.2-2.

2. Environmental Monitoring and Support Laboratory, Handbook for Analytical Quality Control in Water and Wastewater Laboratories, EPA-600/4-79-019, U.S. EPA, Office of Research and Development, Cincinnati, 1979, p.10-1.

3. Gray, C. S., "Total Quality Control in Japan - Less Inspection, Lower Cost," Business Week, No. 2697, July 20, 1981, pp.23-44.

4. Feigenbaum, A. V., Total Quality Control, McGraw-Hill, New York, 1961.

5. Juran, J. M. and F. M. Gryna, Quality Planning and Analysis, McGraw-Hill, New York, 1970.

6. Quality Costs Technical Committee, "Guide for Reducing Quality Costs," American Society for Quality Control, Milwaukee, 1977.

7. Quality Costs - Cost Effectiveness Committee, "Quality Costs - What and How," 2nd edition, American Society for Quality Control, Milwaukee, 1971.

8. Grant, E. L., and R. S. Leavenworth, Statistical Quality Control, 4th edition, McGraw-Hill, New York, 1972.

9. Duncan, A. J., Quality Control and Industrial Statistics, 4th edition, Richard D. Irwin, Inc., Homewood, Illinois, 1974.

10. Wernimont, G., "Statistical Control of the Measurement Process." In: Validation of the Measurement Process, ACS Symposium Series No. 63, American Chemical Society, Washington, D.C., 1977, pp.1-29.

11. Mandel, J., "Measurement and Statistics," Quality Progress, 14(8), 1981, pp.34-36.

12. Eisenhart, C., "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems." In: Precision Measurement and Calibration, NBS Special Publication 300, U.S. Department of Commerce, National Bureau of Standards, 1969, pp.21-47.

13. Murphy, R. B., "On the Meaning of Precision and Accuracy." In: Precision Measurement and Calibration, NBS Special Publication 300, U.S. Department of Commerce, National Bureau of Standards, 1969, p.358.

14. American Society for Testing and Materials, "Standard Practice for Determination of Precision and Bias of Methods of Committee D-19 on Water," ASTM Designation: D2777-77. In: 1977 Annual Book of ASTM Standards, Part 31, pp.7-19.

15. Massart, D. L., A. Dijkstra and L. Kaufman, Evaluation and Optimization of Laboratory Methods and Analytical Procedures, Elsevier Scientific Publishing Co., New York, 1978.

16. Dodge, H. F., "Keep It Simple," Journal of Quality Technology, 9(3), 1977, p.102.

17. Hamaker, H. C., "Seeing Myself as a Shewhart Medalist," Quality Progress, 14(1), 1981, pp.24-27.

SECTION 3

QA/QC TOOLS

Commonly used QA/QC tools are discussed in this section in terms of:

- Purposes and potential benefits
- Information required to judge effectiveness
- Qualitative guidance for effective use
- Sources of further information

Among the possible reasons for using a quality control tool are documentation, appraisal, control or improvement of quality, and prevention of quality problems. Some tools serve more than one of these purposes. The methods available are not all suitable for every job (1).

For the sake of brevity, some well-documented activities, though important, are not discussed in this section. These include (with references for further information):

- Chain-of-custody procedures (2 - 5)
- Data handling and reporting (2, 5 - 8)
- Organization (1, 9 - 12)
- Preventive maintenance (2, 12 - 13)
- Training (1 - 2)

These activities all serve to prevent quality problems from occurring.

The topics that are covered include:

- Blanks
- Calibration
- Control charts
- Interlaboratory studies
- Material controls
- Method development
- Performance and system audits
- Reference materials
- Replication
- Sampling
- Spike-recovery studies
- Study planning
- Surrogate compounds
- Validation

A brief subsection is devoted to each topic. USEPA's QA audit category identifiers for measured values are given where appropriate. References are listed at the end of each subsection for convenience.

References

1. Feigenbaum, A. V., Total Quality Control, McGraw-Hill, New York, 1961.

2. Environmental Monitoring and Support Laboratory, Handbook for Analytical Quality Control in Water and Wastewater Laboratories, EPA-600/4-79-019, U.S. EPA, Office of Research and Development, Cincinnati, 1979.

3. Environmental Monitoring and Support Laboratory, "Standard Operating Procedures: Chain of Custody," EMSL-CI/1005, U.S. EPA, Office of Research and Development, Cincinnati, 1980.

4. Office of Water Enforcement, NPDES Compliance Sampling Inspection Manual, PB81-153215, U.S. EPA, Washington, D.C., May, 1979.

5. Frank, R. S., "Records - Why Keep Them?" In: Optimizing Chemical Laboratory Performance through the Application of Quality Assurance Principles, Association of Official Analytical Chemists, Arlington, VA, 1980, pp.129-154.

6. MacDougall, D., et al., "Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry," Analytical Chemistry, 52(14), 1980, pp.2242-2249.

7. Currie, L. A. and J. R. DeVoe, "Systematic Error in Chemical Analysis." In: Validation of the Measurement Process, ACS Symposium Series 63, American Chemical Society, Washington, D.C., 1977, pp. 126-130.

8. Ku, H. H., "Expression of Imprecision, Systematic Error, and Uncertainty Associated with a Reported Value." In: Precision Measurement and Calibration, NBS Special Publication 300, U.S. Department of Commerce, National Bureau of Standards, 1969, pp.73-78.

9. Quality Assurance Management Staff, "Interim Guidelines and Specifications for Preparing Quality Assurance Project Plans," QAMS-005/80, U.S. EPA, Office of Research and Development, Office of Monitoring Systems and Technical Support, Washington, D.C., 1980.

10. Wening, R. J., "The Role of the Quality Control Manual in the Inspection and Testing Laboratory." In: Testing Laboratory Performance: Evaluation and Accreditation, NBS Publication 591, U.S. Department of Commerce, National Bureau of Standards, 1980, pp.99-107.

11. Massart, D. L., A. Dijkstra and L. Kaufman, Evaluation and Optimization of Laboratory Methods and Analytical Procedures, Elsevier Scientific Publishing Co., New York, 1978.

12. Juran, J. M. and F. M. Gryna, Quality Planning and Analysis, McGraw-Hill, New York, 1970.

13. Environmental Monitoring and Support Laboratory, "Standard Operating Procedures: Facilities and Equipment," EMSL-CI/1008, U.S. EPA, Office of Research and Development, Cincinnati, 1980.

BLANKS

Blanks are an appraisal tool used to check for bias due to contamination. They sometimes are used also to correct statistically for such bias. Two kinds of blanks are commonly employed (e.g., see (1)):

- Laboratory reagent blank (LRB)* - a solution prepared in the laboratory from inert substances and treated exactly as a laboratory sample for the parameter being measured, including all preparations, holding times, and other pre-analysis treatments. Sometimes called method blank.
- Field reagent blank (FRB)* - a solution prepared from inert substances and treated as a field sample in all aspects, including exposure to the sample bottle, holding time, preservatives and other pre-analysis treatments. Sometimes called field blank.

The blank often consists of distilled (reagent) water when water samples are being analyzed (2). Field blanks provide a more comprehensive check for contamination than method blanks because they are exposed to the full sequence of sample-handling procedures.

Effective use of blanks, in terms of the type of blank and frequency of analysis, requires knowledge of the contamination problems present or most likely to occur. Contamination problems can differ in frequency of occurrence, magnitude and stability over time. Following are some different patterns of contamination and examples of their possible causes:

---

*LRB and FRB are the EPA QA audit category identifiers for the measured values of the blanks.

- Constant from sample to sample - may be caused by contamination of solvents, reagents, or glassware
- Differences among batches of samples - may be caused by inconsistencies in practices of different sampling teams
- Random changes from sample to sample - may be caused by inconsistent analytical practices, or lack of method ruggedness.

Observation of patterns and types of contamination can help to identify and eliminate causes (see (3) and (4) for discussions of potential causes).

Qualitative Guidance

1. Differences between blank and sample procedures can cause bias. It generally is necessary to confirm experimentally that identical procedures are unnecessary (2).
2. Blank results, like any analytical results, are subject to analytical error. Therefore, use of blank results to correct sample results can introduce added variation, correlation or bias, depending on the correction procedure used (5).
3. Extrapolating blank results to other samples can be misleading if the contamination process is unstable (differs from sample to sample).
4. The frequency of analysis of blanks can be reduced when experience demonstrates that preventive measures have made contamination unlikely.
5. Analysis of blanks for quality documentation is ineffective when contamination consistently occurs at the same level (apart from analytical error). Once the existence of such contamination is identified, one should seek to eliminate its causes rather than continue to document its existence.

References

1. Environmental Protection Agency, "Guidelines Establishing Test Procedures for the Analysis of Pollutants," Federal Register, 44(233), December 3, 1979, pp.69464-69575.

2. Wilson, A. L., "Performance Characteristics of Analytical Methods-IV, Talanta, 21, 1974, pp.1109-1121.

3. Murphy, T. J., "The Role of the Analytical Blank in Accurate Trace Analysis." In: Accuracy in Trace Analysis, Vol. 1, NBS Special Publication 422, U.S. Department of Commerce, National Bureau of Standards, 1976, pp.509-539.

4. Zief, M. and J. W. Mitchell, Contamination Control in Trace Element Analysis, Wiley, New York, 1976.

5. Wilson, A. L., "The Performance Characteristics of Analytical Methods-II," Talanta, 17, 1970, pp.31-44.

CALIBRATION

The calibration function for an analytical method is the mathematical relationship between the analytical response (Y) and the sample concentration (X). The simplest (and most desirable (1)) relationship is $Y = bX$, where b is the calibration constant or response factor. A number of questions arise concerning the calibration function, namely:

- What is its mathematical form (e.g., linear)?
- What is the range of validity (e.g., linearity)?
- How is it best estimated experimentally?
- How stable is the relationship?

General discussions of calibration problems can be found in references (2) to (5). Two alternative calibration schemes are offered in many analytical methods:

1. External standard calibration, and
2. Internal standard calibration.

Both the external and internal calibration procedures in each of the methods require the preparation of calibration standards at multiple concentration levels for each parameter. The levels should bracket the expected range of concentrations in samples to be analyzed.

Statistical methods have been developed for evaluating and ensuring calibration effectiveness. Response surface methods (6) are designed to optimally identify the form of and estimate mathematical relationships.* Regression methods have been used to develop estimating equations for calibration constants in commonly occurring functional forms (see (5) and (7 - 9)). The use of optimum design and estimation procedures can minimize the effect of calibration error on the bias and precision of analytical results.

Statistical estimation of the calibration function can be illustrated with the relationship

$$Y = bX \tag{3.1}$$

defined above. The object of calibration is to estimate the response factor, b. Suppose that n calibration samples are analyzed, and that the result on the ith sample, which has known concentration $X_i (>0)$, is Y_i. Suppose that the standard deviation of Y is proportional to the concentration (as is the case for many chemical methods). That is, $\mathrm{Var}(y) = X^2\sigma^2$. Then the weighted regression estimator of the response factor is

$$\hat{b} = \frac{1}{n} \sum_{i=1}^{n} R_i = \bar{R} \tag{3.2}$$

---

*The form and range of validity of the calibration function should be identified during method development.

with $R_i = Y_i/X_i$(5). R_i if often called a response factor and thus $\bar{R}$, the weighted regression estimator, is the mean response factor. Calibration effectiveness can be evaluated using the estimated variance of b, s^2/n, where

$$s^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (R_i - \bar{R})^2 \tag{3.3}$$

However, there is another way of evaluating calibration effectiveness that is more directly related to end-use quality. The calibration function is used to estimate the concentration of routine samples by observing Y and computing

$$X = Y/b. \tag{3.4}$$

It can be shown (10) that the standard deviation of a concentration estimated in this manner is approximately

$$\text{Std. Dev. } (\hat{X}) = \left[\frac{X^2\sigma^2}{b^2}\left(1 + \frac{1}{n}\right)\right]^{\frac{1}{2}} \tag{3.5}$$

This formula, with b and σ^2 replaced by their estimates $\hat{b}$ and s^2, can be used to judge the impact of calibration error on analytical precision (e.g., the effect of the number of calibration samples, n).

For example, if $\hat{b} = 10$ based on 3 standard injections, and $\sigma =$ 15% for repeated injections, then

$$\text{Std. Dev. } (\hat{X}) = \left[\frac{X^2\,(0.15)^2}{10^2}\left(1 + \frac{1}{3}\right)\right]^{\frac{1}{2}} = 0.017X;$$

i.e., the standard deviation of concentrations estimated using $\hat{b}$ based on 3 concentrations will be about 1.7% of the estimated concentration.

Estimating methods are also available for other commonly occurring functional forms of the calibration curve (see reference (6) for further information). Using the wrong functional form can be a significant source of analytical bias. Two alternate forms that should be considered are linear with non-zero intercept and non-linear curves. Natrella (5) describes procedures to evaluate if those alternative models are appropriate and gives estimating formulae for each functional form.

Another important consideration is the frequency of calibration. Many analytical methods require verification of the calibration curve each day that analyses are done using one or more calibration standards. If the response for any parameter differs by more than x percent from the predicted response, the test must be repeated with a fresh standard, or a new calibration curve prepared. Variation in calibration constants can be a significant source of analytical variation and too infrequent recalibration can result in analytical bias. Conversely, too frequent recalibration of a stable system can inflate analytical precision and result in less accurate data. To determine an effective calibration schedule, the pattern of variation in calibration constants (response factors) must be studied.

Control charts (see next section) can be effective in monitoring the variation in calibration constants or response factors. Historical data on the variation in response factors over time can be used to develop a cost-effective strategy for calibration checks and recalibration.

Qualitative Guidance

1. Use of the wrong functional form in calibration (e.g., using zero intercept when the intercept is nonzero) can be a serious source or bias (11 - 12).

2. Differences between calibration sample and regular sample can cause bias. For example, the calibration procedures recommended in some methods do not include the sample preparation, clean-up, or extraction phases of the methods. Wilson (4) discusses the possible problems which can arise when calibration standards are treated differently than samples in an analytical method. Biases in the steps that are skipped in the calibration process will not be corrected for by the calibration curve. Additional quality control procedures like spiked samples and surrogates are required to evaluate potential biases in the steps not included in the calibration process.

3. At least three different concentrations covering the range of interest should be used in initial calibrations to allow a check of whether the calibration function is linear. If the function is found to be consistently linear, the minimum number of calibration standards required to bracket the expected concentrations in samples to be analyzed should be used.

4. The order of analysis of calibration samples should be randomized to avoid biasing estimates of calibration constants (3).

5. Variation in calibration constants (e.g., changes in b in (3.1) over time) can be a serious source of analytical error (13). Calibration results can be monitored to check stability (e.g., using control charts).

6. Too infrequent recalibration of an unstable system can result in analytical bias. However, an unstable system is best dealt with by identifying and eliminating causes of instability (14). One cause of unstable calibration is unstable calibration standards (11).

7. Too frequent recalibration of a stable system can damage precision. More frequent recalibration does not guarantee better quality (15). Determining an effective calibration schedule requires characterizing the pattern of variation in response factors (e.g., by estimating variance components for response factors - within day, between day, etc.).

8 Calibration data is an important potential source of information on analytical precision (16). However, repeated readings on portions of the same sample may not give a realistic evaluation of calibration accuracy (repeated readings do not reflect differences due to matrix effects, for example (17)).

9. Weighted least squares (regression) methods for estimating calibration constants are preferable to "eyeball" methods since their effectiveness can be quantified (5). (Note that regression estimates for common calibration functions can be computed from simple formulae as illustrated above).

10. Extrapolation of the calibration function beyond the standard concentration used in developing the function can lead to biases. This is true at very low concentrations (where background interference may play an important role) as well as at high concentrations (where linearity may not hold).

References

1. Lashoff, T. W., "The Measuring Process and Laboratory Evaluation." In: Testing Laboratory Performance, NBS Special Publication 591, U.S. Department of Commerce, National Bureau of Standards, 1980, pp.25-30.

2. Juran, J. M., (ed.), Quality Control Handbook, 3rd edition, McGraw-Hill, New York, 1974.

3. Wilson, A. L., "The Performance Characteristics of Analytical Methods-II," Talanta, 17, 1970, pp.31-44.

4. Wilson, A. L., "The Performance Characteristics of Analytical Methods-IV," Talanta, 21, 1974, pp.1109-1121.

5. Natrella, M. G., "Characterizing Linear Relationships Between Two Variables." In: Precision Measurement and Calibration, NBS Special Publication 300, U.S. Department of Commerce, National Bureau of Standards, 1969, pp.204-249.

6. Myers, R. H., Response Surface Methodology, Allyn and Bacon, Boston, 1971.

7. Hunter, J. S., "Calibration and the Straight Line: Current Statistical Practices." Journal of Association of Official Analytical Chemists, 64 (3), pp. 574-583, 1981.

8. Williams, E. J., "Regression Methods in Calibration Problems." Bulletin International Statistical Institute, 43, pp. 17-28, 1969.

9. Rosenblatt, J. R. and Spiegelman, C. H., Discussion of "A Baysian Analysis of the Linear Calibration Problem" by W. G. Hunter and W. F. Lamboy. Technometrics, 23(4), pp. 329-333, 1981.

10. Kendall, M. G. and A. Stuart, The Advanced Theory of Statistics, Volume 1, 3rd edition, Griffin, London, 1969, p.232.

11. Cardone, M. J. and P. J. Palermo, "Potential Error in Single-Point-Ratio Calibrations Based on Linear Calibration Curves with a Significant Intercept," Analytical Chemistry, 52(8), 1980, pp.1187-1191.

12. Jonckheere, J. A. and A. P. Deheenbeer, "Statistical Evaluation of Calibration Curve Nonlinearity in Isotope Dilution Gas Chromatography/Mass Spectrometry," Analytical Chemistry, 55(1), 1983, pp. 153-155.

13. Sauter, D., C. Kieda, R. Devine and H. Norwicki, "Quantitative Determination of Priority Pollutants - Gas Chromatography-Mass Spectrometry Response Factor Variation." In: Measurement of Organic Pollutants in Water and Wastewater, ASTM STP 686, American Society for Testing and Materials, 1979, pp.221-233.

14. Garden, J. S., D. G. Mitchell, and W. N. Mills, "Nonconstant Variance Regression Techniques for Calibration-Curve-Based Analysis," Analytical Chemistry, 52(14), 1980, pp.2310-2315.

15. Greb, D. J., "Calibration Intervals Specification and Instrument Quality," Journal of Quality Technology, 11(1), 1979, pp.88-94.

16. Ku, H. H., "Expressions of Imprecision, Systematic Error, and Uncertainty Associated with a Reported Value." In: Precision Measurement and Calibration, NBS Special Publication 300, U.S. Department of Commerce, National Bureau of Standards, 1969, pp.73-78.

17. Linnig, F. J. and J. Mandel, "Which Measure of Precision?" Analytical Chemistry, 36(13), 1964, pp.25A-32A.

CONTROL CHARTS

Control charts are graphical methods for monitoring and improving analytical quality (e.g., bias and precision) over time. They are process-control tools that can be applied to spiked sample or reference material recoveries, calibration constants, or other QC test results. They can be used to document data quality, detect the existence of quality problems (special causes), motivate better performance and improve the analytical process. Control charts are most useful when the same analysis is being performed on many samples over time (1). General information on control charts is given in (2) and (3). Information on the application of control charts to analytical work is contained in references (4) to (7).

The following information is required to make effective use of control charts:

- The pattern of statistical variation expected in a process (needed to choose testing frequency) (8)
- The number of parameters measured by a method (multivariate parameter reduction procedures or multivariate control charts should be considered if the number of parameters for each sample is too large to effectively monitor)
- Process control targets (e.g., the average percent recovery of a method)
- The approximate probability distributions of QC test statistics (needed to set realistic control limits) (9) and (10).

The patterns of variation in a process may vary among laboratories, so this information must be developed in individual laboratories. Distributional models and benchmark quality information should be produced by method developers. Quality targets should be based on end-use needs.

Different kinds of control charts are available for controlling different aspects of analytical quality (e.g., bias or precision). The key element of any chart, however, is the control limits that indicate the magnitude of random variation that can be expected to occur when quality objectives are being met or when the analytical process is stable.

One of the most commonly used control charts is the $\bar{X}$ ("X-bar") chart for controlling the process average. The use of this chart requires periodic performance of n QC tests (e.g., determination of n spiked-sample recoveries). The average of each set (subgroup) of test results,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \qquad (3.6)$$

is plotted on a chart of the form illustrated in Figure 3-1. The center-line on the chart indicates the QC target, μ_o (e.g., the desired average percent recovery). The upper and lower control limits (UCL and LCL) are given by

$$UCL_{\bar{X}} = \mu_o + 3\sigma/n^{\frac{1}{2}} \qquad (3.7)$$

and

$$LCL_{\bar{X}} = \mu_o - 3\sigma/n^{\frac{1}{2}} \qquad (3.8)$$

where σ is the short-term standard deviation of the process. The control limits are set so that the chance of $\bar{X}$ falling outside these limits is small when the process mean equals the target value. When the process mean is different from the target, the chance of $\bar{X}$ falling outside the control limits increases with increasing deviations from the target. Thus a point outside the control limits is taken as evidence that the process is "out-of-control"; these special causes of quality problems are investigated and corrected if they can be found.

X-Bar charts with n=1 are commonly used in laboratories to monitor percent recovery, blank concentrations, and calibration parameter estimates.

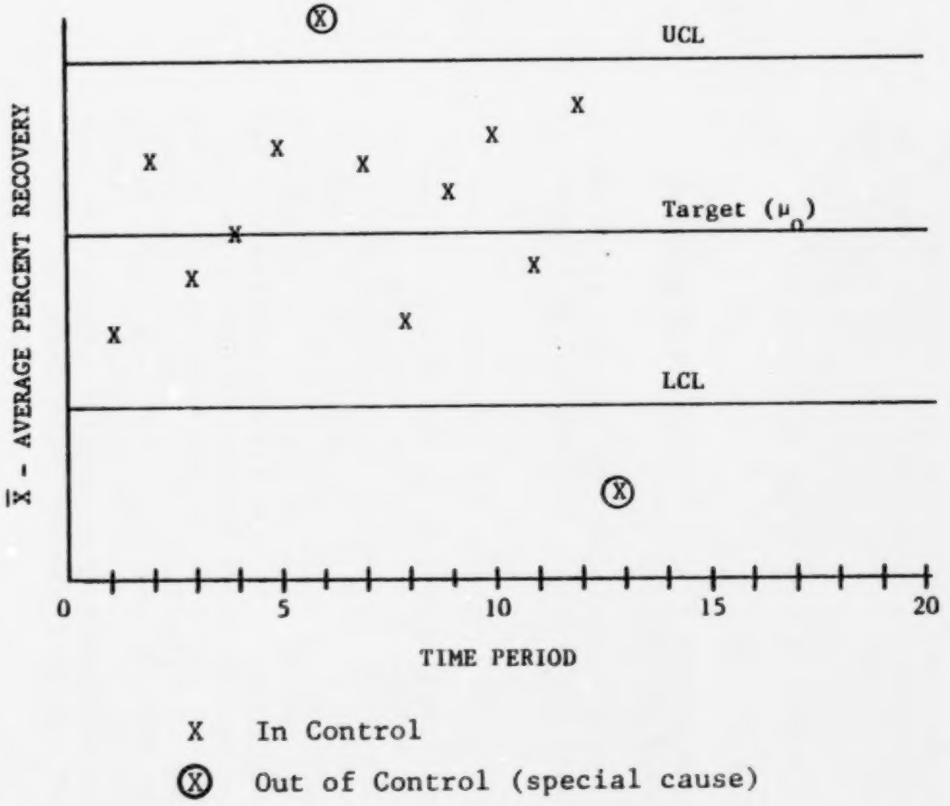


Figure 3-1. $\bar{X}$ Control Chart Illustration

The following are two forms of control charts commonly used to check analytical precision by means of duplicate analyses:

- R charts- based on the range,

  $$R = |X_1 - X_2|, \tag{3.9}$$

  of duplicate analyses. These charts are appropriate when precision is independent of concentration or when only a narrow range of concentrations is of interest (as may be the case in compliance monitoring for NPDES permits). The upper control limit for an R chart based on duplicate analysis is

  $$UCL_R = 3.69\sigma, \tag{3.10}$$

  where σ is the desired short-term process standard deviation (3).

- RSD charts - based on the percent relative standard deviation (coefficient of variation),

  $$RSD = 100R/\sqrt{2}\,\bar{X}, \tag{3.11}$$

  of duplicate analyses (R and $\bar{X}$ are defined above). This chart can be used when the process standard deviation is proportional to concentration and it is necessary to analyze samples with a wide range of concentrations (as may be the case in contract or research laboratories). An approximate method for determining the upper control limit of an RSD chart as a function of the desired RSD is described in reference (11). One alternative to the RSD control chart is an R chart on the logarithms of duplicate analyses (12). Additional information on control charts for precision can be found in references (2) and (3).

Other types of control charts which are useful in laboratory applications include the Difference Control Chart (13) and the Cumulative Sum Chart (14). These charts are more complicated to set up and use than the X-Bar and R charts, but they may be more effective in identifying special causes in some cases.

All of the control charts discussed above apply to a single quality characteristic (e.g., the average percent recovery of a single compound). When a method measures several constituents of each sample, as do all the 600 series methods, one QC approach is to keep a separate control chart for each compound measured. But this approach has the following shortcomings:

- It requires a great number of control charts for some methods. Table 3-1 shows the numbers of compounds measured by 1979 versions of USEPA's 600 Series GC and GCMS methods for pollutant analysis. Much greater numbers of compounds are now of interest in ground-water contamination.
- The risk of violating the control limits of at least one control chart when all targets are met increases with the number of charts kept. Thus, the chance of receiving false out-of-control signals can be large for methods that measure many parameters. The effect of the number of analytes on the false out-of-control error rate is illustrated in the following table, which assumes that analytical errors are uncorrelated and that the out-of-control error rate for each single analyte is 5%.

Number of Analytes	Sample-wise Error Rate (%)
2	9.8
5	22.6
10	40.1
25	72.3
50	92.3
100	99.4

For example, if one measures 100 compounds with uncorrelated analytical errors, one is practically certain to get at least one out-of-control signal when the alpha-level for each compound is 5%. The results in the table represent the worst case, as one would

TABLE 3-1. NUMBERS OF COMPOUNDS COVERED BY USEPA WASTEWATER METHODS

Number of Compounds	EPA Method Numbers*
2	603,605
3	607
4	609
5	611
6	606
7	602
9	612
11	604,625(A)
16	610
25	608,625(P)
29	601
30	624
47	625(B/N)

*A, P and B/N indicate the acid, pesticides and base/neutral fractions of Method 625.

not expect analytical errors to be uncorrelated. Information on the intercompound correlation structure for multi-compound analytical methods is not readily available.

- Separate charts can give conflicting signals as to whether the measurement process is in control.

These problems show the need for some form of data summarization before applying control charts to multi-compound test methods.

One method of taking the number of analytes into account (based on Bonferron's inequality) merely widens control limits on individual charts so that the sample-wise error rate has the desired value (or less). This is done for the $\bar{X}$ chart simply by using the Z-value corresponding to α/p, where α is the desired sample-wise chance of a false out-of-control signal and p is the number of analytes. Table 3-2 gives appropriate Z-values for selected numbers of analytes. The drawback of this approach is evident in the way Z increases with p for given α, thereby decreasing the power of the individual $\bar{X}$ chart. This method is incorporated by EPA for wastewater analysis using Methods 1624 and 1625 (15). These GC/MS methods typically involve the routine analyses of all the analytes (30 or more) included in the scope of the methods.

Another approach to manage the sample-wise error rate, and thus summarize false out-of-control signals, is to retest for analytes that are out-of-control before taking further action. If the second test result is in control, then take no further action. The probability of a false out-of-control signal for a particular analyte thus becomes α^2 where α is the chance of a false out-of-control signal for a single analyte. The power of the control chart to detect special causes is also reduced with this approach. EPA (15) uses this approach for Wastewater Methods 601-613, 624, and 625. These methods include analyses of a large

TABLE 3-2. BONFERRONI Z-VALUES FOR MULTIPLE TESTS

Number of Analytes	Sample-Wise Chance of False Out-of-Control Signal (α in percent)			
	5	2.5	1	0.5
1	1.645	1.960	2.326	2.576
2	1.960	2.241	2.576	2.807
5	2.326	2.576	2.878	3.090
10	2.576	2.807	3.090	3.291
25	2.878	3.090	3.353	3.540
50	3.090	3.291	3.540	3.719

number of analytes, but in typical use, only a subset of the analytes are actually monitored.

A third method of controlling an analytical process with a large number of analytes is a multivariate control chart. The basic idea of a multivariate control chart is to summarize the quality information in measurements on several parameters into a single statistic that can be plotted on one control chart. The multivariate statistic takes interparameter correlations into account, thereby increasing sensitivity to departures from quality targets. Because the multivariate chart is based on a single statistic, its control limits can easily be set to provide the desired protection against false out-of-control signals. The multivariate chart saves paperwork and chart analysis efforts when the process is in control (i.e., most of the time in a well-run laboratory). When it gives an out-of-control signal, results for individual compounds can be scrutinized, if necessary, to identify the quality problem. Overall, then, the multivariate control chart provides a cost-effective way to interpret complex QC data.

Multivariate quality control methods have been used to some extent in the manufacturing and processing industries for many years (16). The universal availability of computers has removed the primary obstacle to more widespread use.

The multivariate control chart called the X^2 (chi-squared) chart can be used under the same conditions as the $\bar{X}$ chart. The only additional information needed is the correlations (or covariances) between tests on different compounds in the same sample. These can be estimated from historical QC data from an in-control measurement process (e.g., using data from old $\bar{X}$ charts, excluding out-of-control points). Suppose that a test method measures p parameters on each QC sample, obtaining results $X_1, \ldots, X_p$. Let v_{ii} be the variance (square of the standard deviation) for the i^{th} compound, and let v_{ij} be the covariance between tests on the i^{th} and j^{th} compounds. Then the matrix of variances and covariances of $X_1, \ldots, X_p$ is $V = [v_{ij}]$. If we denote the matrix inverse of V by $V^{-1} = [v^{ij}]$, and we assume that V is known (estimated from a large amount of data), then the statistic to use for a multivariate control chart is

$$\chi^2 = n \sum_{i=1}^{p} \sum_{j=1}^{p} (\bar{X}_i - R'_i)(\bar{X}_j - R'_j)v^{ij} \quad (3.12)$$

where n = number of results averaged,
p = number of compounds tested,
R'_i = QC target for compound i,
$\bar{X}_i$ = average result for that compound, and
v^{ij} = (i,j)th element of V^{-1}.

To compute this statistic requires inverting the matrix V (once, before starting the χ^2 chart); this can be done with readily available computer software.

In routine application, Equation 3.12 shows that computation of the χ^2 statistic requires only arithmetic operations. Deviations from target for any compound contribute to the value of the statistic. The value of χ^2 is plotted on a single chart. The main difference in appearance compared to the $\bar{X}$ chart is that the control limits are not equidistant from the centerline. The χ^2 control limits are obtained from tables of the chi-squared distribution with p degrees of freedom.

Table 3-3 gives UCL, LCL and centerline values for different p's of interest for USEPA 600 series methods. For example, for Method 606 (phthalate esters), which measures p = 6 compounds on each sample, Table 3-3 gives UCL = 14.4, LCL = 1.24 and centerline 5.35. The tabled control limits were chosen to given 5% probability of a false out-of-control signal.

TABLE 3-3. PARAMETERS FOR χ^2 CONTROL CHARTS

Number of Compounds (p)	USEPA Method Number	Control Chart Parameters		
		LCL	Centerline	UCL
2	603,605	.051	1.39	7.38
3	607	.216	2.37	9.35
4	609	.484	3.36	11.1
5	611	.831	4.35	12.8
6	606	1.24	5.35	14.4
7	602	1.69	6.35	16.0
8		2.18	7.34	17.5
9	612	2.70	8.34	19.0
10		3.25	9.34	20.5
11	604,625(A)	4.40	10.3	21.9
16	610	6.91	15.3	28.8
25	608,625(P)	13.1	24.3	40.6
29	601	16.0	28.3	45.7
30	624	16.8	29.3	47.0
47	625(B/N)	30.0	46.3	67.8

Note: LCL, Centerline and UCL are the 2.5, 50 and 97.5 percentiles of the χ^2 distribution with p degrees of freedom. A, P and B/N indicate the acid, pesticide and base/neutral fractions of Method 625.

As with the $\bar{X}$ chart, when variances (and covariances) are estimated from just a few sample results, control limits must be computed from a different distribution. In the multivariate case the distribution to use is Hotelling's T^2 distribution.

To illustrate the application of a X^2 chart to wastewater data, consider USEPA Method 605 for benzidines (17). This method measures p = 2 compounds, benzidine and 3,3'-dichlorobenzidine, by high performance liquid chromatography with electrochemical detection. The interlaboratory study for the method (18) showed average percent recoveries of 63 and 67 percent. An analysis of the recoveries from distilled water reported in that study showed a between-compound correlation of about 0.7. Now if we assume that a laboratory has relative standard deviations 25 and 30 percent for the two compounds, then the variance-covariance matrix is

$$V = \begin{bmatrix} .0248 & .0222 \\ .0222 & .0404 \end{bmatrix}$$

(e.g., v_{11} $(.63 \times .25)^2 = .0248$ and $v_{12} = .63 \times .25 \times .67 \times .30 \times .7 = .0222$), and the inverse of V is

$$V^{-1} = \begin{bmatrix} 79.4 & -43.6 \\ -43.6 & 48.7 \end{bmatrix}$$

If we use the average percent recoveries from the interlaboratory study for QC targets and analyze n = 1 QC sample per day, control chart values can be obtained from the formula

$$X^2 = 79.4(X_1 - .63)^2 + 2(-43.6)\ (X_1 - .63)(X_2 - .67) + 48.7\ (X_2 - .67)^2 \qquad (3.13)$$

(by Equation 3.12).

Table 3-4 shows twenty simulated recoveries for benzidine and 3,3'-dichlorobenzidine. All results were generated using relative standard deviations of 25 and 30 percent and a correlation of 0.7. The following mean recoveries were used in the simulation: samples 1 to 5 and 11 to 15, 63 and 67 percent; samples 6 to 10, 63 and 27 percent; and samples 16 to 20, 40 and 80 percent. Thus, the simulation illustrates results when the measurement process is on target, when it is off target for one compound, and when it is off target in opposite directions for the two compounds. The X^2 statistic, calculated using equation 3.13, is shown in Table 3-4 for each sample. To compute the X^2 value for sample 1, for example, substitute 0.72 and 0.60 into formula 3.13 for X_1 and X_2 to get 1.[illegible].

Figure 3-2 shows $\bar{X}$-charts for each compound and a multivariate control chart. Control limits for $\bar{X}$-charts from equations (3.7) and (3.8) are 32 and 94 percent for benzidine and 38 and 106 percent for 3,3'-dichlorobenzidine. The centerline and control limits for the multivariate chart are from Table 3-3 with $p = 2$.

There were two out-of-control periods in the simulated process. In the first, represented by samples 6 to 10, mean recovery for 3,3'-dichlorobenzidine was 27% (40% below the target). The $\bar{X}$-chart for 3,3'-dichlorobenzidine shows two results below the LCL during this period (samples 7 and 8). The multivariate chart shows four results above the UCL during the same period (samples 6,7,9 and 10).

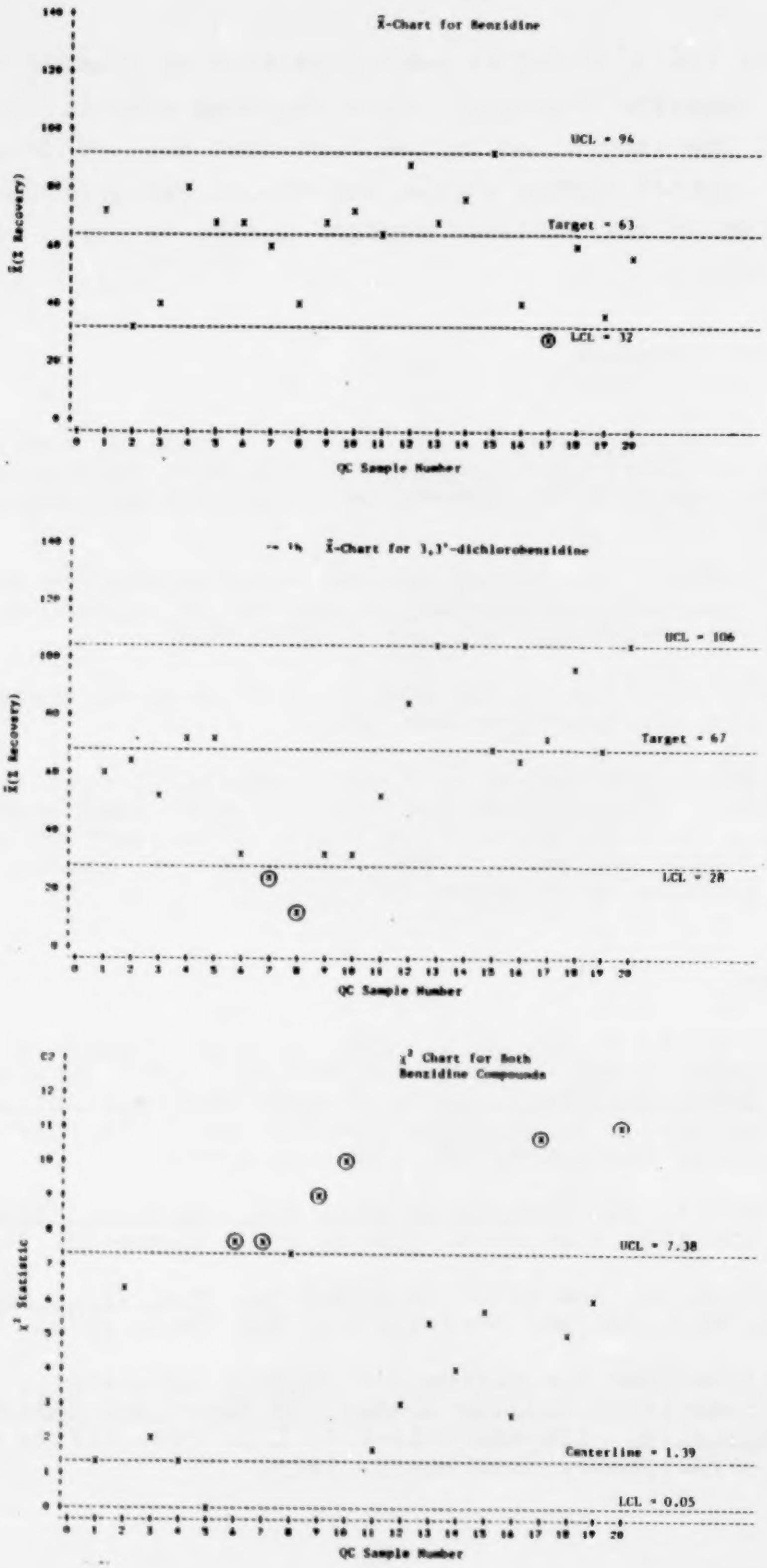
The other out-of-control period is represented by samples 16 to 20. Mean recoveries for benzidine and 3,3'-dichlorobenzidine were 40 and 80 percent during this period (both off target, one below and one above). The $\bar{X}$ chart for benzidine shows one point below the LCL in this episode (sample 17). The X^2 chart has two points above the UCL in the same period (samples 17 and 20).

TABLE 3-4. SIMULATED PERCENT RECOVERIES FOR EXAMPLE

Sample Number	Recovery for Benzidine*	Recovery for 3,3'-Dichlorobenzidine*	χ^2 Statistic
1	72	60	1.3
2	33	64	6.5
3	41	53	2.0
4	78	73	1.2
5	67	72	0.1
6	70	34	7.6
7	60	25	7.7
8	41	14	7.4
9	69	30	9.0
10	73	31	10.0
11	65	51	1.6
12	89	82	2.9
13	66	103	5.2
14	75	104	3.9
15	90	68	5.8
16	42	62	2.8
17	28	71	10.8
18	60	97	5.1
19	37	70	6.1
20	54	106	11.0

$$\text{*Recovery} = \frac{\text{measured concentration}}{\text{prepared concentration}} \times 100$$

Figure 3-2. Multivariate Chart Example



In general the χ^2 chart is more sensitive to changes in quality than the separate $\bar{X}$-charts. This improved sensitivity is evidenced in the simulation by the fact that more χ^2 points are outside the control limits during out-of-control episodes. Further information on multivariate control charts is given in References (19) to (21).

Qualitative Guidance

1. The simplicity and visual impact of control charts can be lost if charts are required on too many parameters. There is no use generating more data than can be handled effectively (22).
2. The reasons for out-of-control results must be identified, corrected and documented if control charts are to be an effective analytical process control tool.
3. Timely plotting of results and follow-up is needed for effective corrective action (23).
4. Control limits based on a small amount of data (used to estimate σ) "may differ greatly from what they should be" (24). As a consequence, a larger than expected proportion of results may fall outside these control limits even when the process is in control.

References

1. McCully, K. A. and J. G. Lee, "Quality Assurance of Sample Analysis in the Chemical Laboratory." In: Optimizing Chemical Laboratory Performance Through the Application of Quality Assurance Principles, Association of Official Analytical Chemists, Arlington, VA, 1980, pp.57-86.
2. Duncan, A. J., Quality Control and Industrial Statistics, 4th edition, Richard D. Irwin, Inc., Homewood, IL, 1974.
3. Grant, E. L. and R. S. Leavenworth, Statisical Quality Control, 4th edition, McGraw-Hill, New York, 1972.
4. Environmental Monitoring and Support Laboratory, Handbook for Analytical Quality Control in Water and Wastewater Laboratories, EPA-600/4-79-019, U.S. EPA, Office of Research and Development, Cincinnati, 1979.

5. Wernimont, G., "Use of Control Charts in the Analytical Laboratory," Industrial and Engineering Chemistry, 18(10), 1946, pp.587-592.

6. Bennett, C. A. and N. L. Franklin, Statistical Analysis in Chemistry and the Chemical Industry, Wiley, New York, 1954.

7. Eisenhart, C., "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems." In: Precision Measurement and Calibration, NBS Special Publication 300, U.S. Department of Commerce, National Bureau of Standards, 1969, pp.21-47.

8. Wernimont, G., "Statistical Control of the Measurement Process." In: Validation of the Measurement Process, ACS Symposium Series No. 63, American Chemical Society, Washington, D.C., 1977, pp.1-29.

9. Moore, P. G., "Normality in Quality Control Charts," Applied Statistics, 6(3), 1957, pp.171-179.

10. Morrison, J., "The Lognormal Distribution in Quality Control," Applied Statistics, 7(3), 1958, pp.160-172.

11. Iglewicz, B. and R. H. Myers, "Comparison of Approximations to the Percentage Points of the Sample Coefficient of Variation," Technometrics, 12(1), 1970, pp.166-170.

12. Environmental Monitoring and Support Laboratory, Quality Assurance Handbook for Air Pollution Measurement Systems, Volume I - Principles, EPA-600/9-76-005, U.S. EPA, Office of Research and Development, Research Triangle Park, NC, 1976, p.22 (Appendix H).

13. Grubbs, F. E., "The Difference Control Chart with an Example of Its Use," Industrial Quality Control, July, 1946, pp.22-25.

14. Page, E. S., "Cumulative Sum Charts," Technometrics, 3(1), 1961, pp.1-9.

15. Environmental Protection Agency "Guidelines Establishing Test Procedures for the Analyses of Pollutants Under the Clean Water Act," Federal Register 49 (209), October 4, 1984, pp. 43234-43406.

16. Jackson, J. E. and R. H. Morris, "An Application of Multivariate Quality Control to Photographic Processing," Journal of the American Statistical Association, 52, 1957, pp. 186-199.

17. Environmental Protection Agency, "Guidelines Establishing Test Procedures for Analysis of Pollutants," Federal Register 44 (223), December 3, 1979, pp. 69464-69575.

18. Kinzer, G., et al., EPA Method Study 15, Method 605, Benzidines, (Draft Final Report), EPA Contract No. 68-03-2624, undated.

19. Elder, R. S. and L. P. Provost. "Efficient Control Charts for Wastewater Laboratories." American Laboratory, 15(7) July 1983, pp. 82-93.

20. Jackson, J. E., "Quality Control Methods for Several Related Variables," Technometrics, 1(4), 1959, pp.359-377.

21. Montgomery, D. C. and H. M. Wadsworth, "Some Techniques for Multivariate Quality Control Applications," ASQC Technical Conference Transactions, 1972.

22. Frazier, R. P., J. A. Miller, J. F. Murray, M. P. Mauzy, D. J. Schaeffer and A. F. Westerhold, "Establishing a Quality Control Program for a State Environmental Laboratory," Water and Sewage Works, 121(5), 1974, pp.54-57.

23. Phillips, R. J. and M. K. Wilson, "On-line Control Charts for Chromatography," American Laboratory (31), 1984, pp. 26-32.

24. Hillier, F. S., "X and R-Chart Control Limits Based on a Small Number of Subgroups," Journal of Quality Technology, 1(1), 1969, pp.17-26.

INTERLABORATORY STUDIES

There are three reasons for conducting interlaboratory studies (1):

- Evaluate a test method (usually the final stage of method development)
- Compare alternative methods
- Evaluate laboratory performance (to ensure compatibility among laboratories).

For general information on interlaboratory studies, see references (1) to (5).

The benefits of interlaboratory studies depend on their purpose. Method evaluation (validation) studies document method performance under a variety of laboratory conditions; they provide information on bias and precision needed for QC and end-use planning. Validation studies sometimes uncover short-comings in test methods. Laboratory evaluation studies identify laboratories with exceptional performance - both good and bad. Eliminating causes of exceptionally bad performance improves interlaboratory precision. Studying laboratories with exceptionally good performance can uncover means of improving the performance of all laboratories. Interlaboratory comparisons can provide motivational and educational benefits (6). They are "one of the most effective elements of a quality assurance plan" (7).

The information needed to plan an effective method validation study includes the likely population of user laboratories, types of samples to which the method will be applied, ranges of concentration of interest, and maximum estimation error tolerable in bias and precision estimates to be derived from the study. Information needed to plan an effective laboratory evaluation program includes the magnitude of differences it is important to detect and the tolerable level of risk of not detecting such differences.

Qualitative Guidance

1. The first concern in planning a method validation study "is to ensure that a workable method exists, as described by a protocol, and that the participating laboratories have achieved a state of statistical control" (Mandel, (3), p.2). Lack of process control is likely to result in long-run variation within laboratories, which in turn results in larger between-laboratory differences.

2. Only when the method protocol is in final form should an interlaboratory study be undertaken to estimate method bias and precision (5).

3. "One practice to be avoided is that of selecting a group of laboratories judged to be those best qualified and equipped for the interlaboratory study." Precision estimates should be obtained under the conditions in which the method will be used in practice (4).

4. Effective method validation requires the use of appropriate statistical methods for design and analysis (2) - (5). The following items should be investigated: magnitudes of relative biases between laboratories, distributional models for recovery variation, relationships of bias and precision to sample concentration, and differential response of laboratories to different concentrations or matrices (interactions).

5. For convenience of use, validation results should be summarized to the greatest extent consistent with statistical analyses. The reporting of test method bias should be preceded by a test for significance (e.g., a statistical test of whether average recovery differs from 100 percent) (5). Analysis of variance (2) can be used to determine the degree of summarization reasonable for a particular compound. Multivariate analysis of variance (8) can be used to determine the degree of summarization reasonable across compounds measured by the same method (e.g., is the average percent recovery the same for all the compounds?). Appropriate use of either of these techniques requires an appropriate statistical model derived from the way the study was conducted.

6. Some observers of validation studies believe that there is too great a tendency to discard outlier results in analyzing validation data (e.g., see (10)). Inappropriate outlier-screening procedures can bias results of statistical analyses. (Most outlier tests assume that results are normally distributed; this is not true for all analytical data.) Carrying out analyses both with and without suspect observations is one way to evaluate the impact of discarding outliers (11).

7. Revisions or refinements in a test method can make results of previous validation studies obsolete (9).

8. In evaluating laboratories one should look for patterns over time or across samples, not overstress results on a single sample. Youden developed methods for analyzing interlaboratory performance data (12).

9. "A coordinating laboratory is ...mandatory for the existence of an interlaboratory quality assurance program" (13).

10. Presenting samples blind to the analyzing laboratory gives the most realistic results (13).

11. Rapid feedback to participating laboratories is important to ensure that evaluations affect performance (14).

12. Laboratories with better than average performance (e.g., nearer 100 percent average recovery) should not be considered better unless they can show why their performance is better (15).

References

1. Nelson, B. N., "Survey and Application of Interlaboratory Testing Techniques," Industrial Quality Control, 23, May 1967, pp.554-559.

2. Youden, W. J. and E. H. Steiner, "Statistical Manual of the Association of Official Analytical Chemists," Arlington, VA, 1975.

3. American Society for Quality Control, Interlaboratory Testing Techniques, Milwaukee, WI, 1978.

4. American Society for Testing and Materials, "Tentative Recommended Practice for Conducting an Interlaboratory Test Program to Determine the Precision of Test Methods," ASTM E-11, undated.

5. American Society for Testing and Materials, "Standard Practice for Determination of Precision and Bias of Methods of Committee D-19 on Water," ASTM Designation: D2777-77. In: 1977 Annual Book of ASTM Standards, Part 31, pp.7-19.

6. Taylor, J. K., "Validation of Environmental Data by Intercalibration and Laboratory Quality Control Programs," Presented before the American Chemical Society, Division of Environmental Chemistry, Los Angeles, CA, 1974.

7. MacDougall, D., et al., "Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry," Analytical Chemistry, 52(14), 1980, pp.2242-2249.

8. Kramer, C. Y. and D. R. Jensen, "Fundamentals of Multivariate Analysis, Part IV," Journal of Quality Technology, 2(1), 1970, pp.32-40.

9. Rhodes, R. C., "Components of Variation in Chemical Analysis." In: Validation of the Measurement Process, ACS Symposium Series No. 63, American Chemical Society, Washington, D.C., 1977, pp.176-198.

10. Byrne, F. P., "The Analyst and Accuracy." In: Accuracy in Trace Analysis, Vol. 1, NBS Special Publication 422, U.S. Department of Commerce, National Bureau of Standards, 1976, pp. 123-126.

11. Kruskal, W. H., "Some Remarks on Wild Observations." In: Precision Measurement and Calibration, NBS Special Publication 300, U.S. Department of Commerce, National Bureau of Standards, 1969, pp.346-348.

12. Youden, W. J., "Ranking Laboratories by Round-Robin Tests." In: Precision Measurement and Calibration, NBS Special Publication 300, U.S. Department of Commerce, National Bureau of Standards, 1969, pp.165-169.

13. Watts, R. R., "Proficiency Testing and Other Aspects of a Comprehensive Quality Assurance Program." In: Optimizing Chemical Laboratory Performance Through the Application of Quality Assurance Principles, Association of Official Analytical Chemists, Arlington, VA, 1980, pp.87-115.

14. Amore, F., "Good Analytical Practices," Analytical Chemistry, 51(11), 1979, pp.1105A-1110A.

15. Youden, W. J., "How to Evaluate Accuracy." In: Precision Measurement and Calibration, NBS Special Publication 300, U.S. Department of Commerce, National Bureau of Standards, 1969, pp.361-364.

MATERIAL CONTROLS

Controls on the quality of materials and supplies used in the laboratory are a major means of preventing quality problems. The general objectives are to purchase materials of satisfactory quality and ensure that their quality does not deteriorate. Methods for achieving these objectives include:

- Specifying quality needs to suppliers
- Testing delivered materials

- Examining supplier (vendor) quality control data
- Optimizing purchasing methods (e.g., using central purchasing (1) or limiting the number of suppliers (2))
- Monitoring the quality of stored materials.

Information on quality needs in terms of purity, stability, etc., should be obtained during method development. General information on material controls can be found in references (3) to (5).

Qualitative Guidance

1. "Beware of changes" in handling procedures, suppliers, batches of materials, etc. (6). Overlapping of changed conditions (e.g., analyzing a given sample using materials from both old and new suppliers) provides protection against quality problems.
2. Documentation of the source, age and quality of materials is an important tool for identifying causes of quality problems.
3. Within-laboratory screening of the quality of delivered materials is not as effective as documented vendor process controls (2).

References

1. Watts, R. R., "Proficiency Testing and Other Aspects of a Comprehensive Quality Assurance Program." In: Optimizing Chemical Laboratory Performance Through the Application of Quality Assurance Principles, Association of Official Analytical Chemists, Arlington, VA, 1980, pp.87-115.
2. Deming, W. E., "What Top Management Must Do," Business Week, No. 2697, July 20, 1981, pp.19-21.
3. Environmental Monitoring and Support Laboratory, Handbook for Analytical Quality Control in Water and Wastewater Laboratories, EPA-600/4-79-019, U.S. EPA, Office of Research and Development, Cincinnati, 1979.

4. Fiegenbaum, A. V., Total Quality Control, McGraw-Hill, New York, 1961.

5. Juran, J. M. and F. M. Gryna, Quality Planning and Analysis, McGraw-Hill, New York, 1970.

6. Rhodes, R. C., "Components of Variation in Chemical Analysis." In: Validation of the Measurement Process, ACS Symposium Series No. 63, American Chemical Society, Washington, D.C., 1977, pp.176-198.

METHOD DEVELOPMENT

According to Deming (1), method development has three stages: obtaining, improving and understanding a response. Improving and understanding a response require experimental investigations to identify and control, at optimum levels, factors with significant effects on quality. Thus ruggedness tests and interlaboratory (validation) studies are important aspects of method development. The latter subject was discussed in the Interlaboratory Studies section. Ruggedness tests are discussed below.

Method validation studies provide information on the following important aspects of quality control (2 - 3):

- Critical steps in the method (with performance tolerances)
- Appropriate calibration procedures
- Expected recovery for the method
- Variance components for the method
- Distributional model for analytical results (for setting control limits)
- Groupings of analytes measured by the method (by similarity of average recovery and precision)

Youden describes the relationship between ruggedness and interlaboratory tests.

Ruggedness tests are experiments that determine whether a measurement process is sensitive to small changes in operating conditions. Among the factors that should be evaluated are:

- Environmental conditions
- Analyst training and experience
- Calibration function (form, stability, and range or validity)
- Interferences and matrix effects (5)
- Materials (purity and stability)
- Options in the method (e.g., calibration procedures, types of equipment, GC columns)

Ruggedness tests identify and set tolerances for critical factors in a method (6); they also identify factors that do not require close control (thus saving unnecessary control efforts). Wernimont (7) describes ruggedness testing methods and cites examples of their application.

The following experimental design for Method 625 follows Youden's recommended approach (8) and is presented as an example of the ruggedness test procedure:

(1) Select seven factors which potentially effect the results of Method 625 with two values of scrutiny for each factor (see Table 3-5).

(2) Obtain enough effluent sample with some compounds present to prepare 17 samples for extraction. Extract one sample to determine background levels. Spike selected base/neutral and acid compounds to achieve two approximate levels for each compound: approximately three times the method detection limit and ten times the detection limit.

TABLE 3-5. FACTORS FOR RUGGEDNESS TEST FOR METHOD 625

	Factor to be Varied	Value 1	Value 2
A	Calibration Procedure	Internal Calibration (A)	External Calibration (a)
B	GC/MS Operator	Operator 1 (B)	Operator 2 (b)
C	Method of Glassware Cleaning	Firing (C)	Chromic Acid Wash (c)
D	Extraction Technique	Separatory Funnel (D)	Continuous (d)
E	Concentration Rate	Fast (E)	Slow (e)
F	pH of First Extraction	11.0 (F)	13.0 (f)
G	Instrument Tuning	High End of Criteria (G)	Low End of Criteria (g)

(3) Extract and analyze the set of eight low level samples and eight high level samples using the design matrix in Table 3-6.

(4) Analyze the set of data for each level of compounds using procedures described by Youden (8) to determine the effect of each factor. Estimate the precision of the method using the study results. If this precision estimate is unsatisfactorily large, the interlaboratory study should not be done until modifications are made to the method. The analysis of the factor effects will suggest where the modifications are needed.

Effective method development is an important means of preventing quality problems from occurring. It provides the cost and quality information needed to design analytical programs that ensure end-use effectiveness.

Qualitative Guidance

1. Options in a method that are not demonstrated experimentally to be equivalent are a potential source of bias (5).
2. Careful attention must be paid to design and analysis of ruggedness tests to obtain required information at a reasonable cost (1, 7 - 8).
3. Minimizing the complexity of the analytical procedure is an effective way to build ruggedness into a method (9).

References

1. Deming, S. N., "Optimization of Experimental Parameters in Chemical Analysis." In: Validation of the Measurement Process, ACS Symposium Series No. 63, American Chemical Society, Washington, D.C., 1977, pp.162-175.

2. Horwitz, W., "Evaluation of Analytical Methods Used for Regulation of Food and Drugs," Analytical Chemistry 54(1), 1982, pp. 67A-76A.

TABLE 3-6. DESIGN MATRIX FOR METHOD 625 RUGGEDNESS TEST

		Value of each Factor (see above for letter codes)						
Test	Factor:	A	B	C	D	E	F	G
1		(A)	(B)	(C)	(D)	(E)	(F)	(G)
2		(A)	(B)	(c)	(D)	(e)	(f)	(g)
3		(A)	(b)	(C)	(d)	(E)	(f)	(g)
4		(A)	(b)	(c)	(d)	(e)	(F)	(G)
5		(a)	(B)	(C)	(d)	(e)	(F)	(g)
6		(a)	(B)	(c)	(d)	(E)	(f)	(G)
7		(a)	(b)	(C)	(D)	(e)	(f)	(G)
8		(a)	(b)	(c)	(D)	(E)	(F)	(g)

3. Kirchmer, C. J., et al, "Factors Affecting the Accuracy of Quantitative Analyses of Priority Pollutants Using GC/MS," *Environmental Science and Technology*, 17(7), 1983, pp. 396-401.

4. Youden, W. J., "Experimental Design and ASTM Committees." In: *Precision Measurement and Calibration*, NBS Special Publication 300, Department of Commerce, National Bureau of Standards, 1969, pp.159-164.

5. Wilson, A. L., "The Performance Characteristics of Analytical Methods-IV," *Talanta*, 21, 1974, pp.1109-1121.

6. Wilson, A. L., "The Performance Characteristics of Analytical Methods-I," *Talanta*, 17, 1970, pp.21-29.

7. Wernimont, G., "Ruggedness Evaluation of Test Procedures." In: *Interlaboratory Testing Techniques*, American Society for Quality Control, Milwaukee, WI, 1978, pp.61-64.

8. Youden, W. J. and E. H. Steiner, "Statistical Manual of the Association of Official Analytical Chemists," Arlington, VA, 1975.

9. American Society for Testing and Materials, "Standard Practice for Determination of Precision and Bias of Methods of Committee D-19 on Water," ASTM Designation: D2777-77. In: *1977 Annual Book of ASTM Standards, Part 31*, pp.7-19.

10. MacDougall, D., et al., "Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry," *Analytical Chemistry*, 52(14), 1980, pp.2242-2249.

PERFORMANCE AND SYSTEM AUDITS

Performance audits are quantitative evaluations of laboratory performance based on the analysis of test samples. System audits are qualitative evaluations of laboratory QA/QC programs. Performance and system audits can be conducted either by an outside agency (e.g., for certification purposes) or by personnel of the laboratory itself (for in-house review of QA/QC effectiveness).

Performance audits may be identical to the laboratory evaluations described in the Interlaboratory Studies section, or they may involve only a single laboratory (e.g., a contractor for a particular project). They are based on quantitative acceptance limits for analytical results (e.g., see reference (1)). Systems audits usually involve: 1) reviews of QA manuals and other evidence of organization, and 2) on-site evaluations to confirm the existence of procedures and facilities and to discuss any shortcomings identified. Checklists are commonly used to facilitate systems audits (examples are given in references (1) to (4)).

Performance and system audits give the user of laboratory services assurance that proper emphasis is placed on producing analytical results of suitable quality. They encourage periodic updating of QA plans and procedures in light of past performance, new knowledge and anticipated needs.

Qualitative Guidance

1. Test samples tend to receive extra attention (unless presented blind), so audit results may reflect a laboratory's capability, not its routine performance level (5).

2. Acceptance limits should take into account the number of parameters to be tested; the chance of failing at least one limit by chance increases with the number of parameters tested.

3. In large-scale performance audits involving many laboratories, obtaining uniform audit samples may be difficult. Variation among audit samples sent to different laboratories should be taken into account when comparing laboratory performance.

4. The use of general checklists may encourage the use of overelaborate QA/QC programs not tailored to end-use needs and thus not cost-effective.

References

1. Colby, B. N., "Development of Acceptance Criteria for the Determination of Organic Pollutants at Medium Concentrations in Soil, Sediments, and Water Samples," EPA Contract No. 68-02-3656, Systems Science and Software, LaJolla, CA, 1981.

2. Bicking, C., S. Olin and P. King, _Procedures for the Evaluation of Environmental Monitoring Laboratories_, Tracor Jitco, Inc., EPA-600/4-78-017, U.S. EPA, Office of Research and Development, Environmental Monitoring and Support Laboratory, Cincinnati, 1978.

3. U.S. Department of the Army, "Quality Assurance Program for U.S. Army Toxic and Hazardous Materials Agency," Aberdeen Proving Ground, MD, August, 1980 (draft).

4. Freeberg, F. E., "Meaningful Quality Assurance Program for the Chemical Laboratory." In: _Optimizing Chemical Laboratory Performance Through the Application of Quality Assurance Principles_, Association of Official Analytical Chemists, Arlington, VA, 1980, pp.13-23.

5. Watts, R. R., "Proficiency Testing and Other Aspects of a Comprehensive Quality Assurance Program." In: _Optimizing Chemical Laboratory Performance through the Application of Quality Assurance Principles_, Association of Official Analytical Chemists, Arlington, VA, 1980, pp.87-115.

REFERENCE MATERIALS

Reference materials are materials with well-characterized properties (concentrations certified by the National Bureau of Standards (NBS), USEPA, etc.) that can be used to maintain accuracy in an individual laboratory or to maintain compatibility among different laboratories.* Uriano and Gravatt (1) discuss the role of reference materials in analytical chemistry. Interesting examples of the use of reference materials are described by Uriano and Cali (2). Use of reference materials has been called "the simplest and most reliable means of checking accuracy (3)." Reference materials are not available for some analytes (4 - 5).

---

*The USEPA QA audit category identifiers for the certified and laboratory measured value of a field reference standard are FRC and FRM.

Qualitative Guidance

1. Inaccurate reference materials can be a serious source of quality problems (6 - 7).
2. A reference material cannot make a poor method good, but it can reveal the method's deficiencies (2).
3. Analysis of reference materials shows the presence, but not the cause, of quality problems. A system for identification and elimination of causes of quality problems is necessary for effective use of reference material results.
4. "Without statistical control of measurement processes in individual laboratories, ...reference materials may be of little value in establishing and maintaining accuracy in multilaboratory networks" (1).
5. For the best test of accuracy, any standard submitted to the laboratory should be disguised so it is not given more care and attention than routine samples (3).
6. When feasible, the purity and homogeneity of reference material should be documented.

References

1. Uriano, G. A. and C. C. Gravatt, "The Role of Reference Materials and Reference Methods in Chemical Analysis," CRC Critical Reviews in Analytical Chemistry, 6(4), 1977, pp.361-411.
2. Uriano, G. A. and J. P. Cali, "Role of Reference Materials and Reference Methods in the Measurement Process." In: Validation of the Measurement Process, ACS Symposium Series No. 63, American Chemical Society, Washington, D.C., 1977, pp.140-161.
3. Skogerboe, R. K. and S. R. Koirtyohann, "Accuracy Assurance in the Analysis of Environmental Samples." In: Accuracy in Trace Analysis, Vol. 1, NBS Special Publication 422, U.S. Department of Commerce, National Bureau of Standards, 1976, pp. 199-210.
4. Josephson, J., "Reference Materials." Environmental Science and Technology, 15(12), pp. 1408-1412, 1981.

5. Alvarez, R., et al, "NBS Standard Reference Materials: Update 1982," Analytical Chemistry 54(12), 1982, pp. 1226A-1243A.

6. Watts, R. R., "Proficiency Testing and Other Aspects of a Comprehensive Quality Assurance Program." In: Optimizing Chemical Laboratory Performance through the Application of Quality Assurance Principles, Association of Official Analytical Chemists, Arlington, VA, 1980, pp.87-115.

7. Horwitz, W. L., R. Kamps and K. W. Boyer, "Quality Assurance in the Analysis of Foods for Trace Constituents," Journal of the Association of Official Analytical Chemists, 63(6), 1980, pp.1344-1354.

REPLICATION

Three purposes of replication are to:

- Estimate the relative contribution of steps in a test method to overall method precision
- Test for changes in the precision of a measurement process over time (e.g., using control charts)
- Improve the precision of estimated concentrations by averaging results of replicate analyses.

Although replication can be done at any stage of the measurement process, the most common forms of replication involve:

- Laboratory replicates - multiple aliquots of the same environmental sample, each of which is treated exactly the same throughout the laboratory analytical procedure.*
- Field replicates - multiple samples taken at the same time and place under identical circumstances, each of which is treated exactly the same throughout the field and laboratory analytical procedures.*

---

*The USEPA QA audit category identifiers for laboratory duplicate results are LD1 and LD2. Additional replicate results are denoted LD3 through LD9. The USEPA QA audit category iden identifiers for field duplicate results are FD1 and FD2.

Other types of replicates occur as modifications of laboratory replicates or field replicates. For example, if duplicate injections of an extracted sample are done, this would not be considered a laboratory replicate since all of the analytical steps were not replicated. The duplicate injections could be used to evaluate the analytical precision attributable to the instrument and quantification steps of the method.

In general, field replicates can be most useful in evaluating variation attributable to the sampling, sub-sampling, handling, and storage aspects of an analysis. But the difference between analytical results for field replicates will also include variation attributable to laboratory factors such as extraction, analysts, reagents, instrumentation, etc. Laboratory replicates which are obtained after the sample is in the laboraory by splitting the sample, will contain sub-sampling variation and variation due to the analytical method. When both field replicates and laboratory replicates are done, the results can be analyzed to estimate the proportion of variation contributed by laboratory and field factors (1 - 3).

General discussions on replication are contained in references (1) to (3). The amount and type of replication required to meet specific quality objectives is discussed in Section 4 of this report.

Qualitative Guidance

1. Obtaining meaningful precision estimates through replication requires that the measurement process be in statistical control (4).
2. The number of replications should always be reported with precision estimates "as should the specific portion of the measurement process to which they apply (5)."

3. The possibility that precision may depend on concentration should be considered in planning a precision study. When random errors in the measurement process are multiplicative, precision tends to be proportional to concentration; in this case, it is convenient to express precision in terms of relative standard deviations. For example, in USEPA Method 624, the analytical result is a product of a response factor and ratios of peak areas to concentrations for the standard and unknown. The result is affected by sample concentration and dilution, which are multiplicative processes, also (6).

4. Replicates must be independent to give useful precision information. Duplicate determinations on the same sample extract made at nearly the same time may not be independent (and they are not affected by the problems of most concern - those caused by changes in performance over time (7 - 8)).

5. A practical difficulty in obtaining useful QC test results is ensuring that replicated samples have nonzero concentrations (9). This issue can be of major importance when planning an environmental study.

6. The best strategy for checking precision may not be the best for improving concentration estimates by averaging.

7. If replicate analyses are subject to different systematic errors, their average may be worse than an individual reading for estimating the true concentration (10).

8. In a measurement process with large bias and good precision, replication is not an effective strategy for improving concentration estimates (11).

References

1. Bennett, C. A. and N. L. Franklin, Statistical Analysis in Chemistry and the Chemical Industry, Wiley, New York, 1954.

2. Rhodes, R. C., "Components of Variation in Chemical Analysis." In: Validation of the Measurement Process, ACS Symposium Series No. 63, American Chemical Society, Washington, D.C. 1977, pp.176-198.

3. Wilson, A. L., "The Performance Characteristics of Analytical Methods-II," Talanta, 17, 1970, pp.31-44.

4. Bicking, C. A., "Precision in the Routine Performance of Standard Tests," ASTM Standardization News, January, 1979, pp.12-14.

5. Merten, D., L. A. Currie, J. Mandel, O. Suschny and G. Wernimont, "Intercomparison, Quality Control and Statistics." In: *Standard Reference Materials and Meaningful Measurements*, NBS Special Publication 408, U.S. Department of Commerce, National Bureau of Standards, 1975, p.805.

6. Janardan, K. G. and D. J. Schaeffer, "Propagation of Random Error in Estimating the Levels of Trace Organics in Environmental Sources," *Analytical Chemistry*, 51(7), 1979, pp.1024-1026.

7. Bicking, C. A., "Inter-Laboratory Round Robins for Determination of Routine Precision of Methods." In: *Testing Laboratory Performance*, NBS Special Publication 591, U.S. Department of Commerce, National Bureau of Standards, 1980, pp.31-34.

8. Wernimont, G., "Use of Control Charts in the Analytical Laboratory," *Industrial and Engineering Chemistry*, 18(10), 1946, pp.587-592.

9. Frazier, R. P., et al., "Establishing a Quality Control Program for a State Environmental Laboratory," *Water and Sewage Works*, 121(5), 1974, pp.54-57.

10. Dorsey, N. E. and C. Eisenhart, "On Absolute Measurement." In: *Precision Measurement and Calibration*, NBS Special Publication 300, U.S. Department of Commerce, National Bureau of Standards, 1969, pp.49-55.

11. Suschny, O. and D. M. Richman, "The Analytical Quality Control Programme of the International Atomic Energy Agency." In: *Standard Reference Materials and Meaningful Measurements*, NBS Special Publication 408, U.S. Department of Commerce, National Bureau of Standards, 1975, pp.75-102.

SAMPLING PROCEDURES

Sampling procedures can play a major role in quality control for chemical analysis. But the laboratory analyst often has little involvement in the sampling process. When an invalid sample is sent to a laboratory, no type or degree of quality control can produce valid analytical results. Potential sampling problems occur during sample selection, reduction or mixing, storage, preservation, or pretreatment. The plan for sample selection is critical in meeting objectives of the sampling/analysis program (Study Planning section).

A sample is a portion of material taken from a larger quantity of material (universe) to represent that universe. Sampling methods range in sophistication from grab sampling to automatic continuous sampling. Both engineering and statistical considerations influence the choice of a sampling procedure. The quality issues in sampling are broad in scope with different critical considerations in each type of application. Detailed information on quality control for sampling procedures is given in references (1) to (8). Three necessary steps in any sampling program for chemical analysis are:

1) description of the universe to be represented by the sample(s),

2) The mechanics of selecting and withdrawing sample material, and

3) The preparation of the laboratory sample from the sampled material.

Qualitative Guidance

1. One important issue in sampling is the decision on whether to combine individual sample increments (compositing) prior to analysis. The issue (often discussed as grab sampling versus continuous sampling) is discussed in references (9) and (10). The following general guidelines can be given:

    - Individual samples should be used when variability or extreme concentration levels is the important issue.

    - The statistical characteristics of a single analysis of a composite sample can be quite different than the arithmetic average of the analysis of each individual sample.

    - Composite sampling can be very cost-effective when average concentration level is the important issue.

2. Sampling bias (due to stratification or seasonal changes in the universe) or variabilty (due to heterogeneity of the universe) can seriously reduce the end-use quality of analytical results.

3. Sample loss or contamination can cause serious systematic errors in concentration estimates, so sample handling practices are as important to end-use quality as procedures for obtaining samples (5).

4. Although it is a distinct problem from analytical variation, the effect of sampling error must be considered in judging the end-use effectiveness of any analytical program. When sampling variation is large, precise control of analytical error alone does not result in high end-use quality.

References

1. American Society for Testing and Materials, Standard Recommended Practice for Sampling Industrial Chemicals, E300-73, ASTM, Philadelphia, 1979.

2. Brumbaugh, M. A., "Principles of Sampling in the Chemical Field," Industrial Quality Control, January, 1954, pp. 6-14.

3. Cochran, W. G., Sampling Techniques, John Wiley & Sons, 1977.

4. Environmental Monitoring and Support Laboratory, Handbook for Analytical Quality Control in Water and Wastewater Laboratories, EPA-600/4-79-019, U.S. EPA, Office of Research and Development, Cincinnati, 1979.

5. Huibregtse, K. R. and J. H. Moser, Handbook for Sampling and Sample Preservation of Water and Wastewater, Envirex, Inc., EPA-600/4-76-049, U.S. EPA, Office of Research and Development, Environmental Monitoring and Support Laboratory, Cincinnati, 1976.

6. Kratochvil, B. and J. K. Taylor, "Sampling for Chemical Analysis," Analytical Chemistry, 53(8), 1981, pp.928A-938A.

7. Krotochril, B. G. and J. K. Taylor, "A Survey of the Recent Literature on Sampling for Chemical Analysis," National Bureau of Standards Technical Note 1153, U.S. Government Printing Office, Washington, 1982.

8. Schweitzer, G. E. and J. A. Santolucito, Editors, Environmental Sampling for Hazardous Wastes, ACS Symposium Series No. 267, Washington, D.C., 1984.

9. Elder, R. S., W O. Thompson, and R. H. Myers, "Properties of Composite Sampling Procedures," Technometrics, 22(2), 1980, pp. 179-186.

10. Schaeffer, D. J., H. W. Kerster, and K. G. Janardan, "Grab Versus Composite Sampling: A Primer for the Manager and Engineer," Environmental Management, Vol. 4, No. 2, 1980, pp. 157-163.

11. Currie, L. A. and J. R. DeVoe, "Systematic Error in Chemical Analysis." In: Validation of the Measurement Process, ACS Symposium Series 63, American Chemical Society, Washington, D.C., 1977, pp.114-139.

SPIKE-RECOVERY STUDIES

The analysis of spiked samples is an appraisal tool that can be used to:

- Determine the bias and precision of a test method (e.g., through interlaboratory studies)
- Determine the accuracy of the measurement process in a particular laboratory
- Test for changes in analytical quality in a laboratory

Appendix B to this report discusses important statistical issues in spike-recovery studies. Important chemical issues include procedures to physically add the spike materials, impact of alternative solvents, the chemical equivalence of the spiked portion of analyte x and the amount of analyte x already in the sample, chemical and solvent interferences, preservation, and holding times for spiked samples. These issues must be specifically addressed for each particular spike-recovery study because they can significantly affect analytical results.

Effective planning of spiking studies requires knowledge of the ranges of concentrations of interest, likely background concentrations, and the relationship of recovery to concentration (if percent recovery is independent of concentration, fewer spike levels may be required).

The percent recovery of the spiked compound is commonly defined by

$$\% \text{ Recovery} = 100 \times \frac{\text{LSF}-\text{LSC}}{\text{LSA}},$$

where

LSO = the measured concentration of the compound in the original environmental sample (prior to spiking),

LSA = the amount by which the concentration of the environmental sample increases due to addition of the pure compound (spiking),

LSF = the measured concentration for the spiked sample

(LSC, LSA and LSF are USEPA's QA audit category identifiers.)

Qualitative Guidance

1. There are several alternative definitions of percent recovery, each with advantages and disadvantages (see Appendix B).

2. When background concentrations are comparable to spike levels, estimates of method bias or precision based on recovery data can make a method look worse than it is (see Appendix B).

3. Statistical uncertainty in recovery data should be considered when deciding whether that data indicate that a method is worse than desired (e.g., has an average recovery different from 100 percent (1)).

4. Since contamination changes statistical properties of percent recovery, blank samples should always be analyzed along with spiked samples to check whether background concentration is truly zero.

5. The use of fixed ("blind") spike levels generally should be avoided. Though it is more convenient for the laboratory, this practice can result in low average spike/background ratios that drastically reduce the power of quality control tests or cause false out of control signals (see Appendix B).

6. Proportional spiking is preferrable when spiking in samples with a possible background. The spike to background ratios should be greater than 1 (Appendix B).

7. In spiking programs for QC testing purposes, continued analysis of samples spiked at the same level can lead some analysts to develop "an attitude of expectation" (2) that limits the usefulness of results. This problem can be avoided by varying spike levels enough to make analyses challenging, yet not enough to affect precision appreciably (when precision depends on concentration).

8. QC samples spiked with concentrations or with combinations of compounds that do not normally occur may not give a realistic reflection of analytical quality. For example, a sample spiked with all the compounds measured by EPA Method 624 may be easily identified by an analyst as a QC sample.

9. Appropriate spiking levels depend on end-use needs. For example, in self-monitoring by NPDES permitees, a spike level near the compliance limit may provide the most relevant QC information.

References

1. American Society for Testing and Materials, "Standard Practice for Determination of Precision and Bias of Methods of Committee D-19 on Water," ASTM Designation: D2777-77. In: 1977 Annual Book of ASTM Standards, Part 31, pp.7-19.

2. Frazier, R. P., et al., "Establishing a Quality Control Program for a State Environmental Laboratory," Water and Sewage Works, 121(5). 1974, pp.54-57.

3. Provost, L. P., and R. S. Elder, "Interpretation of Percent Recovery Data," American Laboratory; 57, December, 1983, pp. 57-63.

STUDY PLANNING

Study planning (experimental design) in analytical programs provides answers to such questions as:

- How many samples should be analyzed?
- How should samples be distributed to laboratories for analysis?
- In what order should samples be analyzed within a laboratory?
- What are the most important sources of error in a laboratory?

General considerations in study planning are discussed in references (1) to (5).

Study planning provides several benefits:

- Obtains the best information for a given cost.
- Ensures that study results will provide answers to questions of interest.
- Permits evaluation of analytical quality while performing analyses for program purposes (e.g., see (5)).
- Enables a laboratory to balance sample workloads.

Principles of study planning can be applied to already-completed studies to evaluate the range of applicability of their results.

The information required for effective study planning includes quantitative study objectives, resources available, factors that can affect responses of interest, measures of analytical quality, and appropriate statistical models.

Qualitative Guidance

1. A common shortcoming of unplanned studies is confusing (confounding) factors so that effects cannot be attributed to specific causes. For example, if different laboratories analyze samples from different treatment facilities, differences in results may reflect either facility or laboratory differences.
2. Study planning for program purposes is primarily a user responsibility, but it requires laboratory input of analytical quality information.
3. Effective study planning requires expertise in the area of application and in applied experimental design. Expertise in both areas is rarely found in a single person.

References

1. Natrella, M. G., Experimental Statistics, NBS Handbook 91, U.S. Department of Commerce, National Bureau of Standards, 1966.
2. Davies, O. L., The Design and Analysis of Industrial Experiments, 2nd edition, Hafner Publishing Co., New York, 1956.
3. Cox, D. R., Planning of Experiments, Wiley, New York, 1958.
4. Box, G. E. P., W. G. Hunter and J. S. Hunter, Statistics for Experimenters, Wiley, New York, 1978.
5. Youden, W. J., "Statistical Aspects of Analytical Determinations," Journal of Quality Technology, 4(1), 1972, pp.45-49.

SURROGATE COMPOUNDS

A surrogate compound is a compound added to an original environmental sample that is not one of the materials found in the sample.* Percent recovery of the surrogate compound is used as an indicator of the quality of results for compounds of interest (1). Surrogates are a potential means of testing the quality of every analytical result. Effective use of surrogates requires appropriate compounds for the problem at hand and sufficient recovery data to set realistic control limits.

---

*The USEPA QA audit category identifier for the amount added is LS2; the identifier for the amount measured is LS1.

The selection of an appropriate surrogate for a particular analyte is a problem for the chemical expert. The validation and quantification of the relationship can be done through appropriate experiments and statistical analysis. Data should be collected over a wide range of analytical performance, purposely allowing analytical "mistakes." Data can be screened for appropriate surrogates using correlation analysis and the relationships quantified using regression analysis (2).

Qualitative Guidance

1. The effectiveness of particular surrogate compounds for detecting problems in analyses of interest should be demonstrated experimentally before their use is required. Trials with in-control measurement processes do not test the ability of surrogates to detect quality problems (since problems are not present).
2. Shortcomings in the purity and stability of surrogates or in spiking procedures can cause misleading results.
3. Surrogate control limits should take into account the number of surrogate compounds employed.

References

1. Environmental Monitoring and Support Laboratory, Handbook for Analytical Quality Control in Water and Wastewater Laboratories, EPA-600/4-79-019, U.S. EPA, Office of Research and Development, Cincinnati, 1979.
2. Draper, N. R. and H. Smith, Applied Regression Analysis, Second Edition, John Wiley and Sons, 1981.

VALIDATION

The term "validation" is used in several senses in the quality control literature. In the sense of making analytical results valid, it can encompass all of quality control (1 - 3). In the sense of determining the validity of analytical results, it also can include a broad range of activities. However, in this section validation is used to describe the process that a laboratory (or analyst) is required to follow to demonstrate the ability to apply a method, before using it to analyze real samples. Examples of such validation procedures are discussed in (4) and (5). Their chief benefit lies in uncovering problems in time to prevent their affecting production samples. Objective validation procedures are based on quantitative decision criteria designed to detect serious problems, yet minimize the occurrence of false signals of trouble.

One aspect of method validation is a ruggedness study (see Section on Method Development). The product of a method validation should be preliminary assessment of the analytical method's bias, precision, sensitivity, specificity, and completeness. Typically a method validation study will include reference standards for analysis and spiking to assess bias, repeated analysis to assess precision, analysis of low-level samples to assess sensitivity, and the analysis of "blank" matrices to assess specificity. The principles of study planning (see Study Planning section) should be used in developing an experimental protocol for method validation.

Validation of a method on a particular matrix may be needed in an individual laboratory if a similar matrix was not included in the method validation study. Comparing method performance on the matrix to its performance on spiked reagent water is a key aspect of such a validation effort. Equivalence of bias and precision on reagent water and sample matrices simplifies laboratory QC, since correction for background is not necessary in analyzing spike recoveries from reagent water.

Qualitative Guidance

1. Test samples must be presented blind to the analyst for realistic evaluation to occur.
2. Acceptance procedures and criteria should be based on statistical principles of experimental design to ensure that problems of interest are likely to be detected.

References

1. Kagel, R. O., "Validation and Priority Pollutant Analysis," Invited Plenary Address, American Chemical Society National Meeting, Division of Environmental Chemistry, San Francisco, 1980.
2. Horwitz, W., "Is Your Analytical System Valid?" Chemtech, March, 1984, pp. 186-191.
3. Taylor, John K., "Validation of Analytical Methods," Analytical Chemistry 55(6), 1983, pp. 601A-608A.
4. Environmental Protection Agency, "Guidelines Establishing Test Procedures for the Analysis of Pollutants," Federal Register, 44(233), December 3, 1979, pp.69464-69575.
5. MacDougall, D., et al., "Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry," Analytical Chemistry, 52(14), 1980, pp.2242-2249.

SECTION 4

MEASURING QA/QC COST EFFECTIVENESS

Measures of analytical quality, such as bias and precision, are useful to the laboratory for evaluating and maintaining its performance. However, since factors in addition to analytical quality often affect the usefulness of results, more comprehensive criteria are needed for measuring end-use effectiveness. The purpose of this section is to present methods for evaluating the cost-effectiveness of particular QA/QC activities from the following viewpoint:

- The effectiveness of a QC procedure cannot be judged without identifying, in terms of quantitative objectives, the reasons for its use
- The quality objectives for a QC procedure should be based on end-use needs
- A cost-effective procedure is one that achieves quality targets at reasonable cost (within available resources and at no greater cost than alternative procedures)
- End-use needs generally are flexible enough to allow an adjustment of QC targets, if necessary, to keep costs reasonable

From this viewpoint, two kinds of tools are needed to develop cost-effective QA/QC programs: 1) means of identifying reasonable QC targets based on end-use needs, and 2) means of evaluating procedures for achieving specified QC targets.

Decision-directing formulae for evaluating QA/QC procedures and choosing targets are presented in the Achieving QC Targets section and End-Use Quality section, respectively. Most of these formulae assume that analytical results are normally distributed. Sometimes analytical measurements, especially trace level analysis, cannot be modeled using the normal distribution. When the normal distribution assumption is inappropriate, the formulae usually can be made applicable through a data transformation (e.g., taking logarithms of analytical results). Information on identifying statistical distributions and choosing data transformations can be found in most applied statistics books (e.g., reference (1)).

ACHIEVING QC TARGETS

In this section it is assumed that QC targets have been specified and that it is necessary to develop effective measures to achieve these targets. The targets should be specified in terms of bias, precision, sensitivity and specificity (as defined in the Measures of Analytical Quality section), and should consist of both desired quality levels and deviations that are considered important to detect when they occur. Accuracy problems (bias or imprecision) generally are handled differently than detection problems (sensitivity or specificity) so these topics are discussed separately in the Achieving Accuracy Targets and Achieving Detection Targets sections. The relationship of laboratory size to the frequency of QC tests is discussed in the Achieving Accuracy Targets section.

Achieving Accuracy Targets

The collection of QC activities called process control (see Process Control section) is aimed at achieving accuracy targets. A key step in process control is detecting serious problems as soon as possible after they occur. Control charts (see Control Charts section) are the most commonly used detection tool. Thus, ensuring the effectiveness of control charts at detecting bias and precision problems is a key to achieving accuracy targets. Two questions that arise in the use of control charts are:

- How many analyses should be included in each subgroup?
- How frequently should control chart tests be made?

Tools for answering these questions are described below.

In order to use QC test results effectively to decide when bias is present, it is helpful to set control limits as described in the Control Chart section. Using "3 sigma" limits on the average ($\bar{X}$) of n readings, the probability of not detecting a bias of size b when analytical readings are normally distributed with standard deviation σ is*

$$P = \Phi[3-\sqrt{n}b/\sigma] - \Phi[-3-\sqrt{n}b/\sigma] \qquad (4.1)$$

The probability of detecting a bias of size b in m independent tests (each based on an average of n readings), therefore, is

$$P_D = 1 - P^m \qquad (4.2)$$

Figure 4-1 (based on End-Use Quality section) shows the probability of detecting bias via an $\bar{X}$-chart, as a function of the amount of bias and the number QC tests included on the chart. Figure 4-1 can be used to answer the following kinds of questions:

1. How many tests are needed in a specified period in order to reliably detect a bias of specified size?

   Suppose it is desirable to detect within a week a change of 20 ppb in the concentration

---

* $\Phi(x) = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{x} e^{-t^2/2}dt$ is the cumulative distribution function of the standard normal distribution. (x) is tabled in most applied statistics books (e.g., (1)).

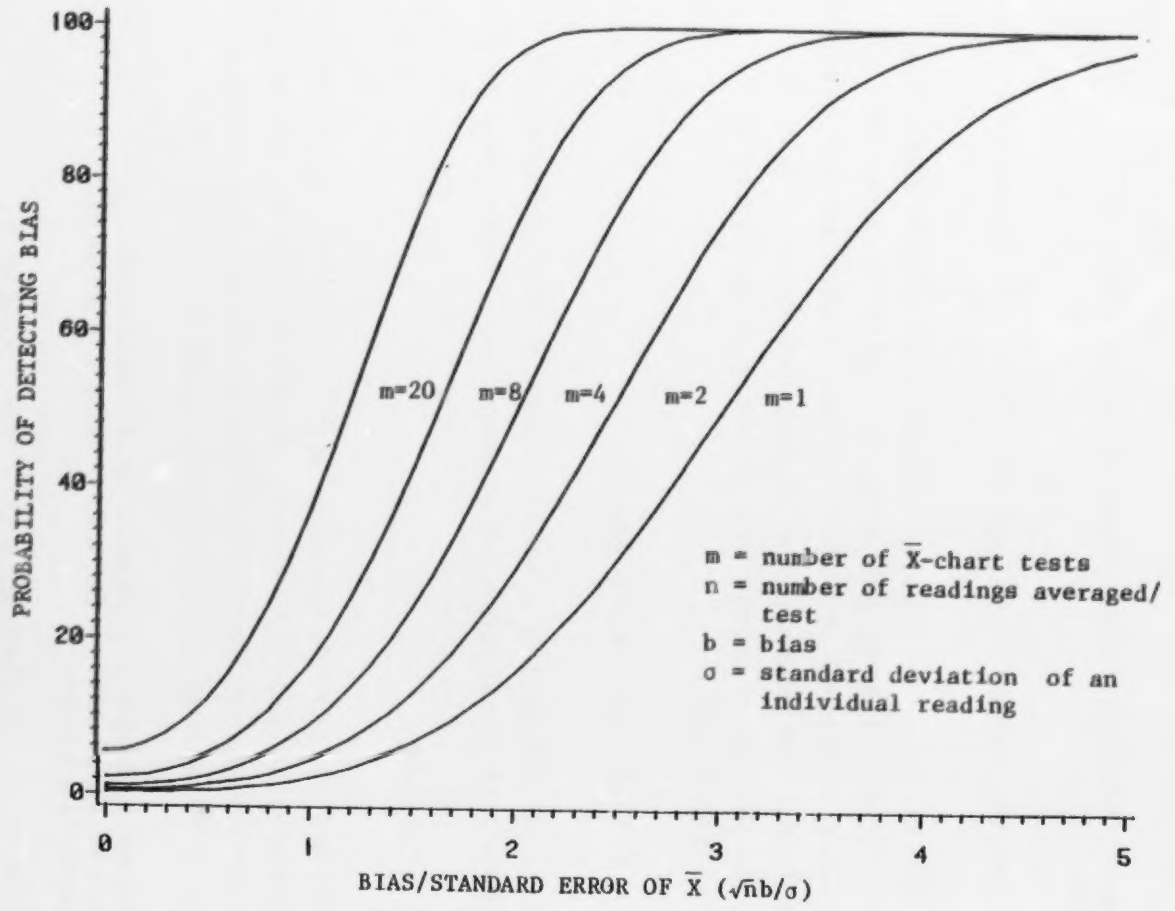


Figure 4-1. Chance of Detecting a Specified Bias in m Points on an $\overline{X}$ Control Chart with 3σ Action Limits

of a calibration standard. If $\sigma=10$ ppb for the analysis of calibration standards and each test is based on duplicate analyses ($n=2$), it is necessary to perform about 5 tests per week to have 95% chance of detecting a bias of 20 ppb. ($20\sqrt{2}/10=2.8$, and $m=5$ gives a 95% chance of detection at this value).

2. What size bias can be reliably detected with a given number of tests?

   Suppose the QC test procedure is based on the analysis of a single spiked sample ($n=1$). Then one test ($m=1$) can detect a bias of 46 percent recovery with 95% probability when $\sigma=10\%$. (For $m=1$, $\sqrt{n}b/\sigma=4.6$ gives a probability of 95%. Solve for b when $n=1$ and $\sigma=10$.)

3. How many replicate analyses are needed for a single test ($m=1$) to reliably detect a specified bias?

   For a 95% chance of detecting a bias of 25 ppb when $\sigma=10$ ppb, $\bar{X}$ should be based on $n=4$ readings. (Solve $\sqrt{n}b/\sigma=4.6$ for n when $b=25$ and $\sigma=10$.)

4. What is the probability of detecting a specified bias in a single QC test ($m=1$)?

   Suppose $b=40\%$ recovery is considered a serious bias and $\sigma=10\%$ recovery is the analytical precision. Then a single spiked-sample recovery ($n=1$) will detect a bias of 40% with probability about 84%. ($\sqrt{n}b/\sigma=4$, and at this value the curve for $m=1$ in Figure 4-1 is at 84%.)

5. What is the probability of detecting a serious contamination problem in the laboratory in a single method blank analysis ($m=n=1$)?

   Suppose a contamination level of 15 ppb is considered serious enough to require corrective action, and experience has shown that it is possible to control contamination at an average level of 5 ppb. If $\sigma=3$ ppb, then a single method blank analysis will detect a shift in contamination from 5 ppb to 15 ppb (a "bias" of 10 ppb) with probability 62%. ($\sqrt{n}b/\sigma=3.3$, and at this value the curve for m=1 in Figure 4-1 is at 62%.).

It can be seen from these examples that Figure 4-1 provides a means of judging the effectiveness of $\bar{X}$-charts applied to many QC measurements aimed at detecting analytical bias (including spiked samples, reference samples, standards, calibration constants and blanks).

The effectiveness of multivariate control charts for detecting bias (see Control Charts section) is more complicated to describe because of the many combinations of biases that can occur for the different analytes of interest. For the most general case, the chance of detecting bias is a function of

$$\lambda = \underline{b}'V^{-1}\underline{b} \qquad (4.3)$$

where $\underline{b}' = (b_1, \ldots, bp)$ is the vector of biases for the p different analytes and V is the matrix of variances and covariances of the analytes. In the case of bias in a single analyte, (4.3) reduces to

$$\lambda = b^2/\sigma^2(1-R^2) \qquad (4.4)$$

where b is the bias, σ is the analytical standard deviation and R is the multiple correlation between results for the biased analyte and the remaining analytes. The chance of detecting a specified bias using a X^2 chart can be obtained from tables of the noncentral X^2 distribution with p degrees of freedom (e.g., reference (2)). The same quantity for a T^2 chart can be evaluated using the noncentral F distribution (e.g., reference (2)) as shown by Anderson (3). In general, increasing λ implies increasing chance of detecting a problem.

The range control chart on duplicate analyses described in the Control Charts section is a common means of checking for changes in analytical precision. The probability of not detecting a change in the analytical standard deviation from the target value (σ_o) to a larger value (σ_1) is given by

$$P = \Pr[\chi^2 \leq 6.81\sigma_o/\sigma_1] \qquad (4.5)$$

The chance of detecting a deterioration in precision in m points on the range chart, therefore, is given by (4.2) with P defined by (4.5).

Figure 4-2 is a graphical representation of (4.5) showing the chance of detecting a given deterioration in precision as a function of the number of R-chart points. It can be used to answer the following questions:

1. How many sets of duplicate analyses are needed in a specified period to reliably detect a specified increase in σ?

   If it is necessary to detect a quadrupling of σ within a week (σ_1/σ_o=4), Figure 4-2 shows that about 6 R-chart points would give a 95% chance of detection.

2. What size change in precision can be reliably detected with a given number of tests?

   A single test (m=1) has only about a 50% chance of detecting a 5-fold increase in σ. With four tests, the chance of detecting a 5-fold increase is about 95%.

The results above for control charts assume that accuracy problems persist at the same level - once they occur - until corrected. If accuracy problems occur intermittently, they are more

---

*P can be found from tables of the cumulative chi-squared distribution with one degree of freedom (since the range of duplicate analyses equals $\sqrt{2}$ times the standard deviation). Formula (4.5) assumes that analytical readings are normally distributed.

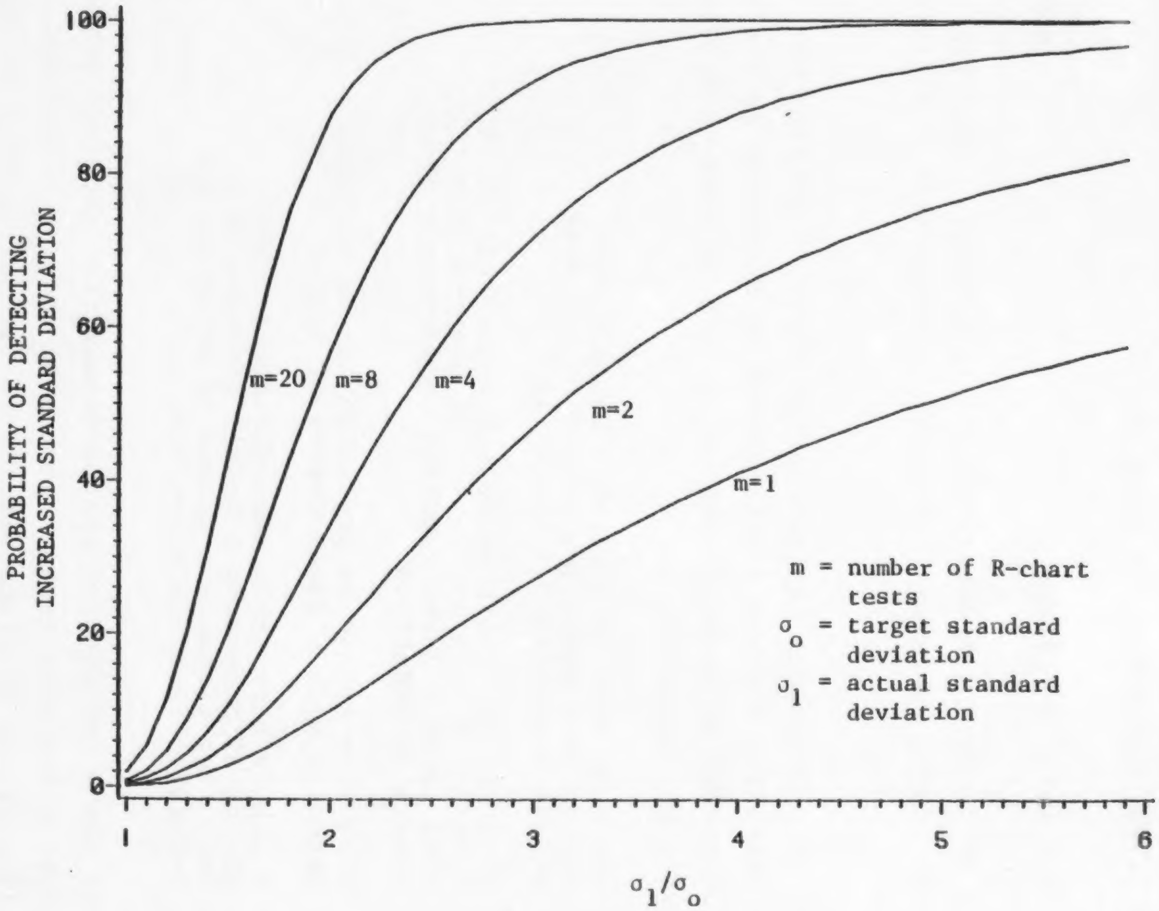


Figure 4-2. Chance of Detecting a Change in Precision in m Points on a Range Control Chart

difficult to detect (but their impact is less serious). Models have been developed to evaluate the effectiveness of control charts when quality problems occur in a random manner (4 - 6), but they are primarily of academic interest (see Simplicity section).

Effective process control requires that problems be corrected once they are discovered. If no effort is made to eliminate persistent problems, much of the effectiveness of control charts will be lost.

Another way of looking at the effectiveness of a QC test is in terms of the average number of tests required to detect a specified problem. This quantity is often referred to as the average run length or ARL of the test. Figure 4-3 shows the ARL as a function of the probability of detecting a problem in a single test. The figure can be related to the earlier results for $\bar{X}$ and R charts by

$$ARL = P/(1-P), \tag{4.6}$$

where P is given by (4.1) or (4.5).

To illustrate the ARL concept, recall that in the fourth $\bar{X}$-chart example, the chance of detecting a bias of 40% recovery on a single test was 0.84. The chance of detecting a bias of 20% in the same example is 0.16. Using (4.6) or Figure 4-3, it can be seen that an average of over five tests would be required to detect this 20% bias. If one test were run per day, this means that an average of one workweek would pass before the bias was discovered. If a test were run every 20 samples, on the average over 100 samples would be analyzed before the problem was detected.

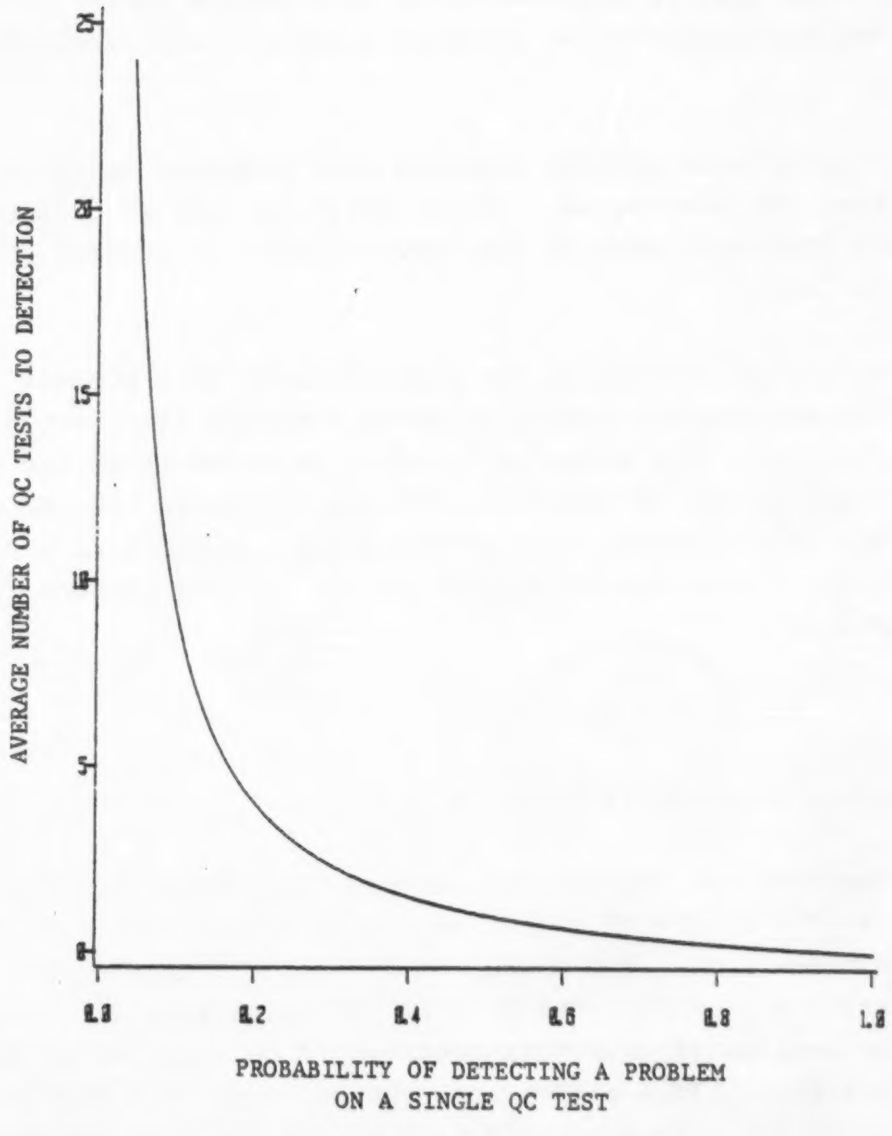


Figure 4-3. Number of QC Tests Required to Detect a Quality Problem

These last examples illustrate another question that arises with many QC activities; namely, should the frequency of QC tests be based on the elapsed time or the number of samples analyzed between tests? The answer depends on the nature of the quality problems likely to occur. For example, if calibration constants tend to remain stable within a day but change between days, QC testing on a day-to-day basis is reasonable. On the other hand, if the occurrence of bias due to GC column deterioration is related to the number of samples analyzed, basing the frequency of QC tests on the number of samples analyzed is reasonable.

Laboratory size can be an important factor in determining the relative costs of quality control when it is appropriate to base QC testing frequency on the elapsed time between tests. For example, if recalibration is required each day analyses are done, the laboratory analyzing ten samples per day will have much smaller calibration costs than the laboratory analyzing two samples per day (on a per sample analyzed basis). On the other hand, the impact of laboratory size is minimal when it is appropriate to base QC testing frequency on the number of samples analyzed between tests. Unfortunately, many quality problems in trace analysis are time-related; e.g., instability of calibration constants and standards, environmental changes and contamination. Thus, small laboratories tend to have a cost disadvantage when it comes to achieving specified QC targets.

In summary, the decision-directing formulae in this section can be used to choose effective means of achieving QC targets for analytical accuracy, but only if the targets are specified in quantitative terms as to desirable and undesirable quality levels. Guidance for relating targets to end-use needs is provided in the End-Use Quality section.

Achieving Detection Targets

Detection problems generally are handled by defining nondetects and detects in a manner that minimizes the rate of occurrence of these problems. Two limits have been defined for this purpose (7):

- Critical level - the level of test result that reliably indicates the presence of a compound
- Detection limit - the true concentration at which a test method reliably detects the presence of a compound.

The critical level protects against false positives (saying a compound is present when it isn't) due to background noise. It should be large enough that samples not containing a compound seldom give that large a test result. Then, when a result exceeds the critical level, one can reasonably conclude that the compound is present. The critical level usually is expressed as a multiple of the standard deviation of background noise (7).

The detection limit addresses the problem of false negatives (saying a compound is not present when it is). The detection limit does not apply to individual test results the way the critical level does: it is a property of a test method that should be used in applying that method. For example, regulatory limits cannot practicably be set below the detection limit (a compound must be detectable at a given level or one cannot check whether it is being held at that level). The detection limit should be reported whenever an analyte is not detected in a sample to indicate the possible level of the analyte (if the analyte is present, its level probably is below the detection limit).

Critical levels and detection limits are commonly determined from distributional models (7 - 8). There are alternative methods less dependent on distributional assumptions; for example, Kagel (9) illustrates one method.* Regardless of what method is used to estimate these limits, the method and conditions of experimentation should be reported along with results.

USEPA's operational definition of the Method Detection Limit is described in reference (10). Further discussions of limits for detection and quantification are in References (11 - 13).

END-USE QUALITY

It was noted in the Definition of Quality as Fitness for Use section that uses of analytical data include estimating concentrations, setting and enforcing regulatory limits, and comparing concentrations from different sources. Measures of effectiveness for each of these uses are described below. The statistical tools discussed in this section can be used to set rational targets for laboratory QC and to ensure effective end-use quality through a comprehensive QA/QC program.

Evaluating Effectiveness in Estimation Problems

The following are examples of problems involving estimation using analytical results:

- Describing treatment system performance
- Establishing a calibration curve
- Setting achievable regulatory limits

---

*Statistical techniques for handling dose-response problems are applicable, also.

- Describing contamination levels in a laboratory
- Documenting average recovery in a laboratory

The answers to these problems are called estimates because they are affected by systematic and random errors in the analytical results (and possibly by detection errors and total failures as well). The purpose of this section is to provide tools for evaluating the impact of analytical errors on estimates of different environmental and analytical parameters of interest.

The following questions arise in estimation problems:

- How effective is a given estimation procedure?
- What estimation procedure gives acceptable results for the least cost?
- What estimation procedure gives the best results for a given cost?

Methods for answering these questions are described below.

One measure of the effectiveness of an estimation procedure is the maximum probable error, the largest error that will occur with specified probability in repeated applications of the procedure.* Estimators often are averages of independent test results. For such estimators we can say approximately that the probability is P that the estimation error is no more than

$$E = z_p \sigma / n^{1/2} \qquad (4.7)$$

where n is the number of results averaged, σ is the appropriate standard deviation or measure of precision of the procedure, and

---

*This criterion only evaluates the impact of random error on estimation. The impact of systematic error can be determined separately if the bias of the measurement process is known.

z_p is the appropriate percentile of the standard normal distribution (11).* For P = 95%, $z_p = 1.96$. This formula should be used in planning estimation studies to ensure that useful results will be obtained. Formula (4.7) can be rearranged to give the number of tests required for an estimator with specified maximum probable error:

$$n = Z_p^2\sigma^2/E^2. \tag{4.8}$$

For example, if a maximum error of 10ppb is desired with 95% confidence and σ = 10ppb, (4.8) indicates that n = 4 tests are needed.

If σ depends on concentration, prior knowledge of concentration may be required to use these formulae. One exception is when the relative standard deviation is constant with respect to concentration;** then the formulae can be applied by interpreting σ and E as relative standard deviation and relative error, respectively. For example, suppose it is desirable to estimate the average recovery in a laboratory with 95% confidence that the estimate will be within 10 percent of the true value if the relative standard deviation is 20 percent. Then (4.8) shows that

$$n = 1.96^2(20)^2/10^2=16$$

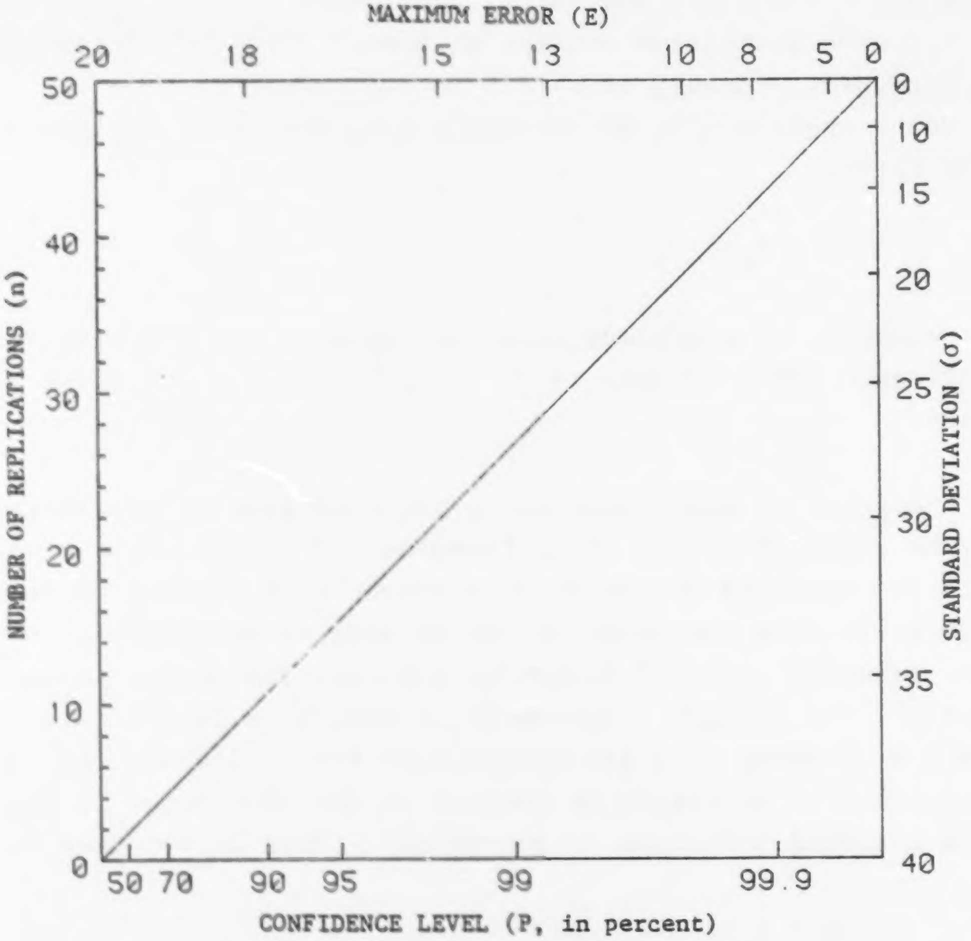
analyses are required.

Formula (4.8) can be evaluated using the nomograph in Figure 4-4. For example, to find the n needed to achieve a maximum error of

---

*The formula is exact if analytical results are normally distributed and σ is known. If results are not normally distributed, the formula improves as an approximation as n increases.

**One case in which RSD is constant with respect to concentration is when analytical results are lognormally distributed.



NOTE: The standard deviation and maximum error must be in the same units (percent, ppb, etc.)

Figure 4-4. Nomograph to Determine the Number of Replications Required to Achieve a Specified Maximum Error

10 ppb, as in the first example above, first find the point where the diagonal intersects the line through E = 10 ppb and P = 95%. Then the line through this point and σ=10 ppb cuts the n scale at the required value, n = 4.

Formula (4.7) also can be evaluated using the nomograph. For example, to determine the maximum probable error that will occur with 95% probability based on n = 4 tests when σ=20 ppb, first find the point where the diagonal and the line through n = 4 and σ=20 intersect, then extend the line through this point and P = 95% to find E = 19.6 ppb.

The n value indicated by (4.8) sometimes will be infeasible for economic reasons. In such cases, the nomograph facilitates finding E and P combinations that yield a practical n. Particular choices of n can be evaluated by finding the diagonal point on the line connecting n and σ, then finding E and P values on lines through this point.

For example, if n = 30 is desirable from a cost standpoint and σ=20 percent, then the following approximate E and P combinations are possible:

E	P
9 percent	99%
8 percent	95%
7 percent	90%
6 percent	75%

It can be seen from the nomograph that n increases with increasing confidence level (P) or decreasing error (E). For fixed n and σ, smaller error requirements mean that a lower confidence level must be accepted.

All factors in the formula except σ can be varied by the user. The standard deviation is characteristic of an in-control measurement process - it can be changed only by changing the process. One valuable result of process control is that it allows one to know σ through experience with the process. Process control gives assurance that the past value of σ is relevant to future analyses.

When σ is unknown, the American Chemical Society's Committee on Environmental Improvement (12) recommends use of the N-N-N rule; that is, run an equal number of field samples, field blanks and spiked blanks. This rule generally will not be cost-effective since it is not tied to either analytical precision or end-use requirements. In most cases some knowledge of σ should be available, if not from studies of the method of interest, then from studies of related methods or from expert opinion (13). One responsibility of method developers is to provide preliminary estimates of important variance components, so that reliance on such sources will not be necessary. Two other approaches available when there is no information on analytical precision are two-stage estimation procedures (14) and pilot studies to estimate the σ values to substitute into formula (4.8). The pilot study approach is less practical because a large number of tests are required to obtain a good standard deviation estimate (over 30 analyses are required to ensure 90% confidence that the error in an estimated standard deviation will be less than 50 percent of the true value (2)).

Cochran (14) describes a simple model for determining a cost-effective sample size (n) when estimation cost is given by

$$C = C_0 + C_1 n + C_2 [b^2 + \sigma^2/n], \qquad (4.9)$$

where

n = number of samples analyzed
σ = analytical standard deviation
b = analytical bias
C_0 = overhead cost
C_1 = cost per sample analyzed
C_2 = cost of estimation error

The formula assumes that the cost of estimation error for an estimate based on n replicate samples is proportional to the mean squared error of estimation (this includes both bias and precision). It can be shown that the value of n that minimizes (4.9) is

$$n = (C_2\sigma^2/C_1)^{1/2} \qquad (4.10)$$

Note that bias does not affect the optimum n (since replication does not reduce bias). The major difficulty with applying this model lies in identifying the cost of estimation error, C_2.

Even if the cost of estimation error cannot be quantified as Cochran's model requires, more effective allocation of resources may be possible (compared to (4.8)) when detailed knowledge of sources of variation is available. Then a replication strategy can be based on variance component and analytical cost information. For example, consider the problem of deciding how many extractions to run on a sample and how many analyses to perform on each extraction. Let

σ_1 = standard deviation due to extraction
σ_2 = standard deviation due to analysis
C_1 = cost/extraction
C_2 = cost/analysis
n_1 = number of extractions/sample
n_2 = number of analyses/extraction.

Then the cost of analyzing a sample is

$$C = n_1C_1 + n_1n_2C_2 \tag{4.11}$$

and the variance of the estimated sample concentration (the average of n_1n_2 analytical results) is

$$\sigma^2 = \sigma_1^{\,2}/n_1 + \sigma_2^{\,2}/n_1n_2 \tag{4.12}$$

Suppose we need to estimate a sample concentration within ± E with confidence P. The most economical allocation of extractions and analyses to meet this requirement is (15)

$$n_2 = (C_1\sigma_2^{\,2}/C_2\sigma_1^{\,2})^{\frac{1}{2}} \tag{4.13}$$

and

$$n_1 = z_P^2(\sigma_1^{\,2} + \sigma_2^{\,2}/n_2)/E^2{*} \tag{4.14}$$

For example, if C_1 = \$48, C_2 = \$20, σ_1 = 10ppb, σ_2 = 15ppb, P = 95% and E = 20ppb, then $n_1 = n_2 = 2$. That is, if two extractions are done and two analyses are run on each extraction, the average of the four analyses will have a maximum error of ± 20ppb with 95% confidence. The cost per sample will be \$176.

If the maximum allowable cost is fixed, then the best n_2 is still determined by (4.13), but n_1 is based on the cost constraint

$$n_1 \leq C/(C_1 + n_2C_2) \tag{4.15}$$

If the maximum allowable cost in the above example is \$90, then $n_2 = 2$ and $n_1 = 1$. Thus one extraction and two analyses would be done on each sample at a cost of $88; the maximum probable error would be ± 29ppb with 95% confidence.

---

*Formula (4.14) assumes that analytical results are normally distributed and σ1 and σ2 are known. It can be used when variance components are proportional to concentration by interpreting the σ's as relative standard deviations and E as relative error.

Figure 4-5 is a nomograph for determining n through formula (4.13). The values of C_2/C_1 and σ_2/σ_1 are computed, then the line through these values gives n_2 on the middle scale. After obtaining n_2, n_1 can be obtained by the other nomograph (Figure 4-4) with $\sigma^2 = \sigma_1^2 + \sigma_2^2/n_2$. To use the nomograph for the above example, read $n_2 = 2$ where the line through $C_2/C_1 = 0.4$ and $\sigma_2/\sigma_1 = 1.5$ crosses the middle scale in Figure 4-5. With $n_2 = 2$, $\sigma = (10^2 + 15^2/2)^{\frac{1}{2}} = 14.6$. Then $n_1 = 2$ can be found using Figure 4-4 with a maximum error of 20ppb and confidence level 95%.

Bennett and Franklin (15) discuss allocation problems of this kind with any number of stages.

Several complications that can occur in the application of this scheme to environmental analyses are beyond the scope of the methods discussed above. These include:

- Correlations between samples or analyses
- Measurement of several parameters on the same sample
- Use of composite sampling procedures in place of arithmetic averages.

Correlations, such as can exist between wastewater samples taken on successive days, make formulas in this section invalid in deciding on the number of days to sample. Examples of methods for dealing with such correlations are given in references (16) to (18). When several parameters are measured on a sample, the methods of this section can be applied separately for each parameter. If results for different parameters conflict, one can pick the result that works best for all parameters or the result

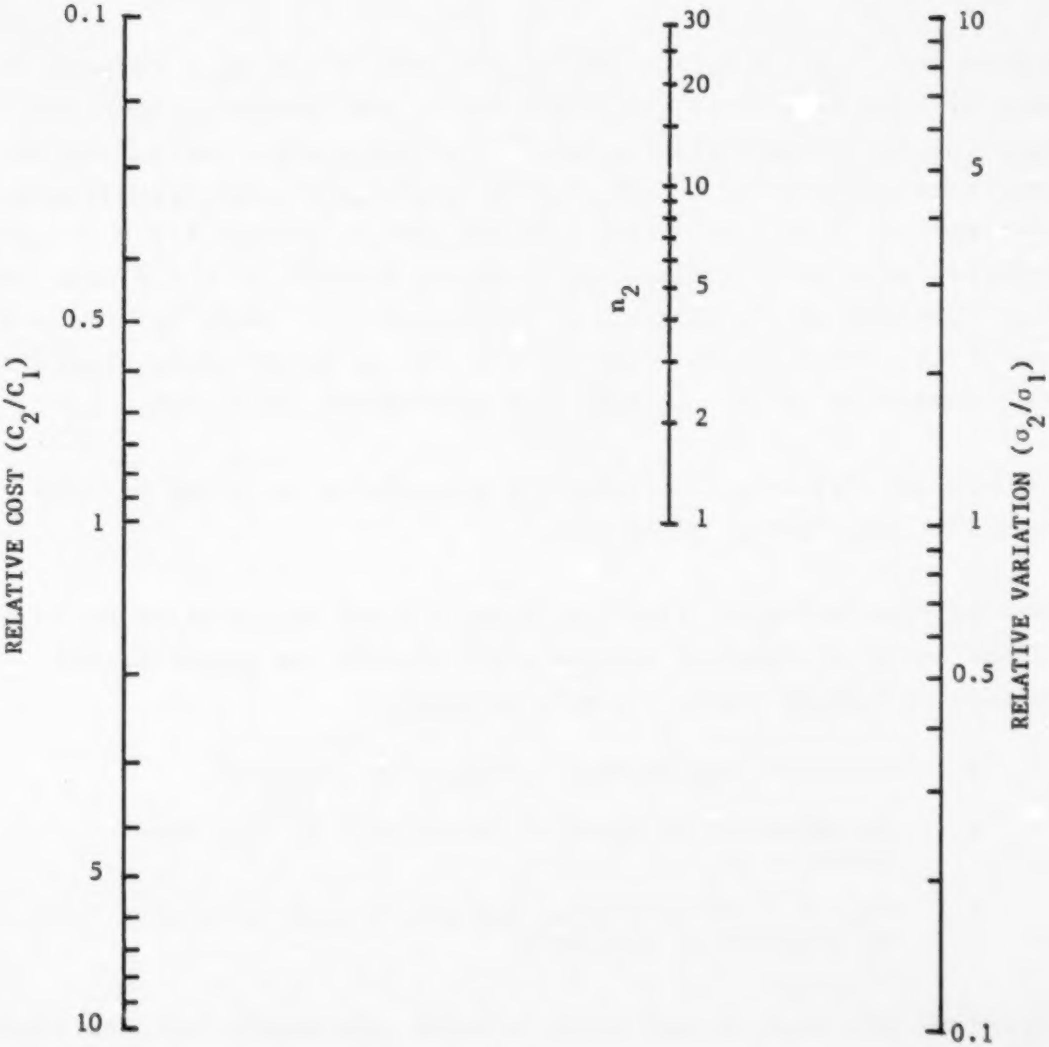


Figure 4-5. Nomograph to Determine the Optimum Number of Replications in the Second Stage of a Two-Stage Procedure

for the most critical parameter (if one exists). Evaluating composite sampling procedures is more difficult than evaluating procedures discussed here (references (19) - (22) give details).

Evaluating Effectiveness in Regulatory Problems

The effectiveness of regulatory programs aimed at limiting water pollution depends on the quality of analytical data available, as well as on the incentives for compliance (23 - 24). The discussion here focuses on detecting violations when they occur. However, unless detecting violations helps induce desired behavior, the measures of effectiveness discussed may be meaningless in terms of accomplishing regulatory objectives.

A common regulatory use of test results is to check whether concentrations of particular compounds exceed regulatory limits. Bias and imprecision in test results can cause two errors in this application: to wrongfully conclude that a violation has occurred, or to wrongfully conclude that one has not occurred. A statistical tool commonly used to show the effectiveness of compliance-testing procedures is the operating characteristic (OC) curve. An OC curve shows the probability of concluding from compliance data that a facility is in compliance, as a function of the true concentration of the compound of interest. OC curves depend on the compliance procedure and on the bias and precision of the sampling and analytical procedures used.*

Figure 4-6 shows OC curves for a procedure that determines compliance by comparing a single analytical result (e.g., the measured concentration for a monthly NPDES self-monitoring sample) to a compliance limit of 100ppb. The two sets of curves show the impact of bias (60 and 90 percent recovery). The three curves

---

*See Duncan (25) for a discussion of OC curves for different compliance procedures.

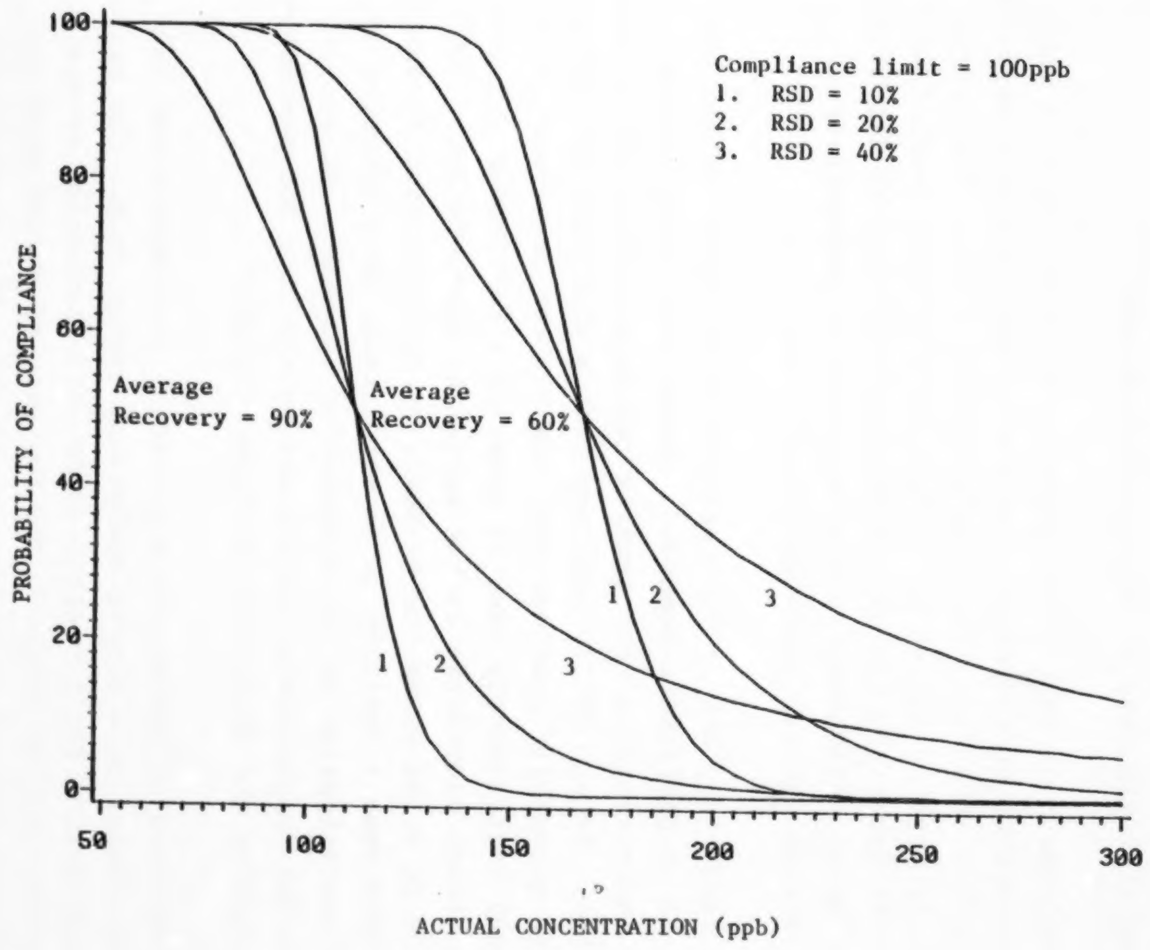


Figure 4-6. OC Curve Example

within each set show the impact of precision (10, 20 and 40 percent RSD). When the true effluent concentration is 150 ppb, the figure shows that the chance of compliance can range from 1 percent (90 percent recovery with RSD = 10 percent) to 87 percent (60 percent recovery with RSD = 10 percent). When the true concentration is 50 ppb, however, the chance of compliance is virtually 100 percent in all cases illustrated. Thus, the impact of analytical quality depends on the relative magnitudes of the true concentration and the compliance limit.

Table 4-1 gives formulae for computing OC curves for three distributional models that occur commonly in analytical QC. The second model was used to produce Figure 4-6.

In general, OC curves show how random and systematic errors in sampling and analysis affect compliance decisions. The magnitude of random variations affects the steepness of the curve; i.e., the ability of the procedure to distinguish between different concentrations. More variation results in less discriminating power, as can be seen by comparing curves for RSD = 10 percent and RSD = 40 percent. The effect of bias is to shift the curve left or right, depending on the direction of bias. (In cases where variability changes with concentration, bias also affects the steepness of the curve.) It can be seen in Figure 4-6 that negative recovery bias shifts the curves to the right; i.e., makes the compliance procedure easier to pass.

Figure 4-7 is an example of an OC curve for two nitrosomine compounds analyzed using USEPA Method 607. The compound recovery and precision values were calculated from data presented in the method development study. The lognormal model from Table 4-1 was used in generating the curves. The compliance limits are hypothetical. The impact of bias (recovery <100 percent) and the effect of laboratory replication on compliance can be seen from these curves.

TABLE 4-1. FORMULAS FOR COMPUTING OC CURVES

Distributional Model	OC Curve Formula*
Normal, constant variance	$Pa = \Phi[(100L - r\mu)/100\sigma]$
Normal, constant RSD	$Pa = \Phi[100(100L - r\mu)/r\mu \mathrm{RSD}]$
Lognormal	$Pa = \Phi[(\log L - \mu')/\sigma']$

* L = compliance limit.
μ = true concentration (same units as L).
σ = standard deviation (same units as L).
r = average percent recovery.
RSD = relative standard deviation ($100\sigma/\mu$).
μ' = $\log [100\mu/r(1+(\mathrm{RSD}/100)^2)^{1/2}]$ (logarithmic scale).
σ' = $[\log (1 + (\mathrm{RSD}/100)^2)]^{1/2}$ (logarithmic scale).
$\Phi(\cdot)$ = cumulative distribution function of standard normal distribution (see Equation 4.1).

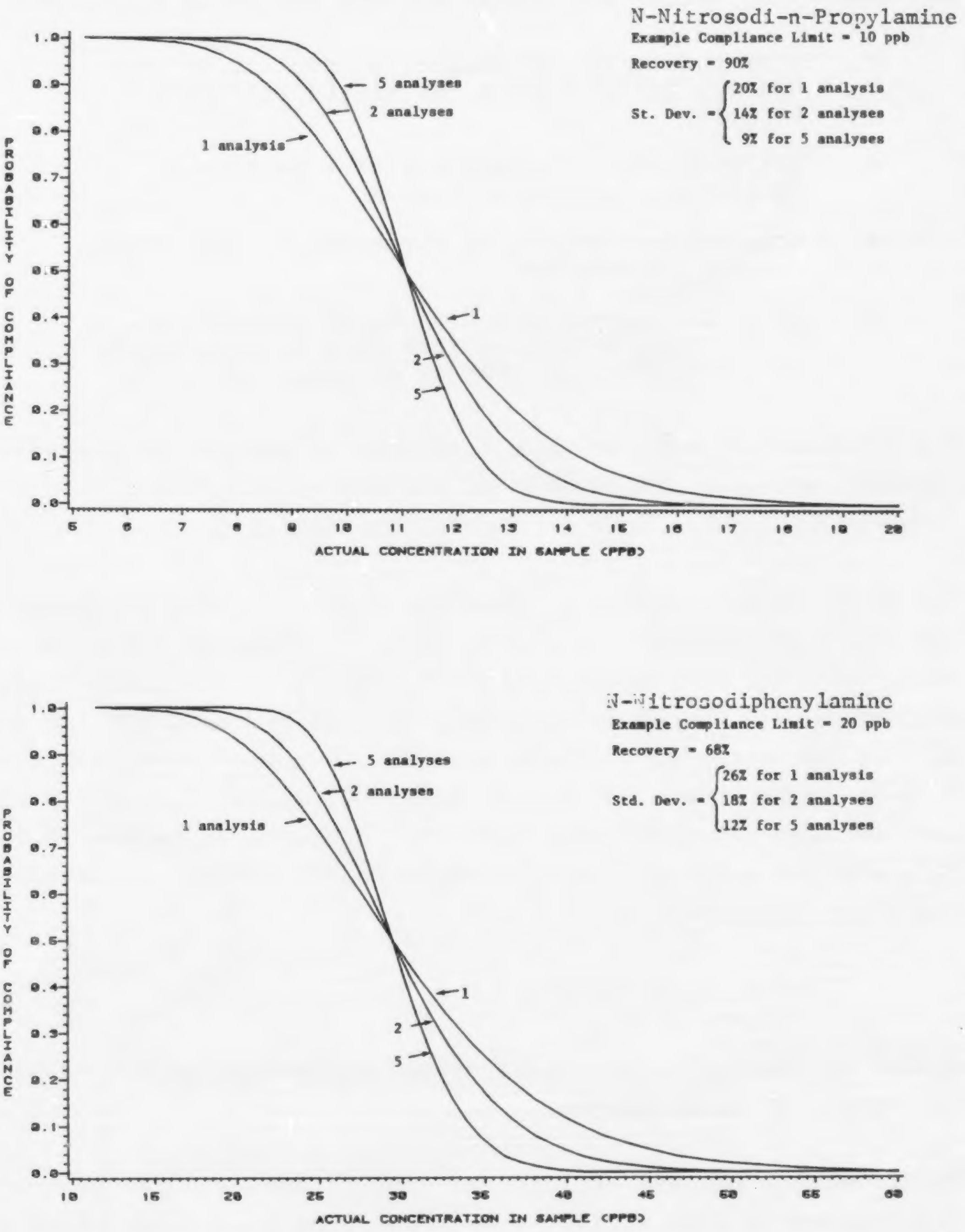


Figure 4-7. Example of Operating Characteristic (OC) Curves for Method 607

In summary, OC curves are useful for the following purposes:

- Show the impact of analytical quality on regulatory effectiveness (an aid to setting QC targets)
- Evaluate the effectiveness of a particular compliance-testing procedure
- Compare performance of alternative compliance-testing procedures
- Show the operational meaning of regulations (e.g., what concentrations must be maintained to ensure minimal risk of noncompliance (26)).

The first use is most obviously related to analytical quality; however, all uses are related to end-use quality and thus are properly of concern in a total quality control program.

The OC curve concept can be combined with cost data to evaluate the cost effectiveness of different QA/QC programs (27). For example, suppose the probabilities of passing a compliance limit *at a given actual concentration above the limit* are P_0 and P_1, for an original and a more costly and effective procedure, respectfully. If the costs of the two alternatives are C_0 and C_1, the cost per violation detected under the original procedure is $C_0/(1-P_0)$, and the *marginal cost* of additional violations detected by the other procedure is

$$MC = (C_0 - C_1)/(P_0 - P_1) \tag{4.16}$$

A plot of marginal cost versus actual concentration can be used to compare QC alternatives.

To illustrate the use of marginal cost to evaluate the cost-effectiveness of a QC activity, consider bias correction based on recoveries from spiked samples and precision improvement via replication. For replication, the probabilities in (4.16) are given (assuming the second model in Table 4-1 holds) by

$$P_1 = \Phi[100\ b(100L - r\mu)/r\mu RSD] \qquad (4.17)$$

where n is the number of replicates. P_O is given by (4.17) with n=1. For recovery correction,

$$P_O = \Phi[100(100(1-d)-r)/rRSD] \qquad (4.18)$$

and

$$P_1 = \Phi[-100d/RSD(1+(1-d)^2/n)^{1/2}] \qquad (4.19)$$

where n is the number of spiked-sample analyses and $d = (\mu - L)/\mu$ (again assuming the second model).* In both cases costs are proportional to the number of samples analyzed, so $C_1 - C_O = n-1$ for replication and $C_1 - C_O = n$ for spiked sample analyses.**

The marginal cost of replication (cost of additional violations detected by replicating laboratory analyses) is illustrated in Figure 4-8 as a function of (coded) actual concentration. The figure shows that the minimum cost per additional violation detected through duplicate analyses (n=2) is about 12 times the cost of detecting a violation based on a single analysis. Marginal costs are even higher when the actual concentration and analytical quality are such that the compliance decision is either clearcut or difficult. For example, suppose the average recovery is 60 percent, the relative standard deviation is 20 percent, the actual concentration is 175 ppb and the compliance

---

*Compliance is determined by comparing $X/(\bar{Y}/L)$ to the compliance limit L; where X is the analytical reading and $\bar{Y}$ is the average recovery of n samples spiked at concentration L.

**The baseline cost (C_O) is assumed to be 1 for convenience. This means that all results here are expressed as relative to C_O.

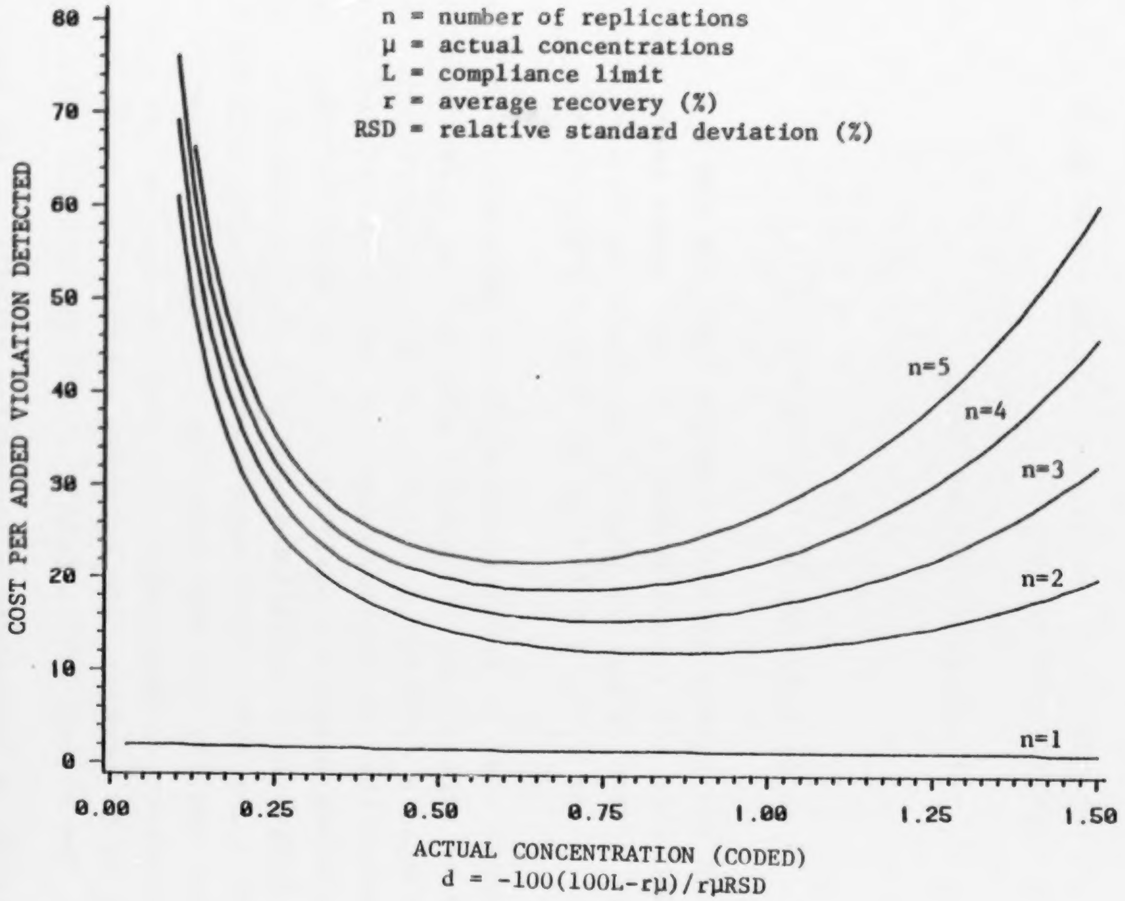


Figure 4-8. Cost Effectiveness of Replication

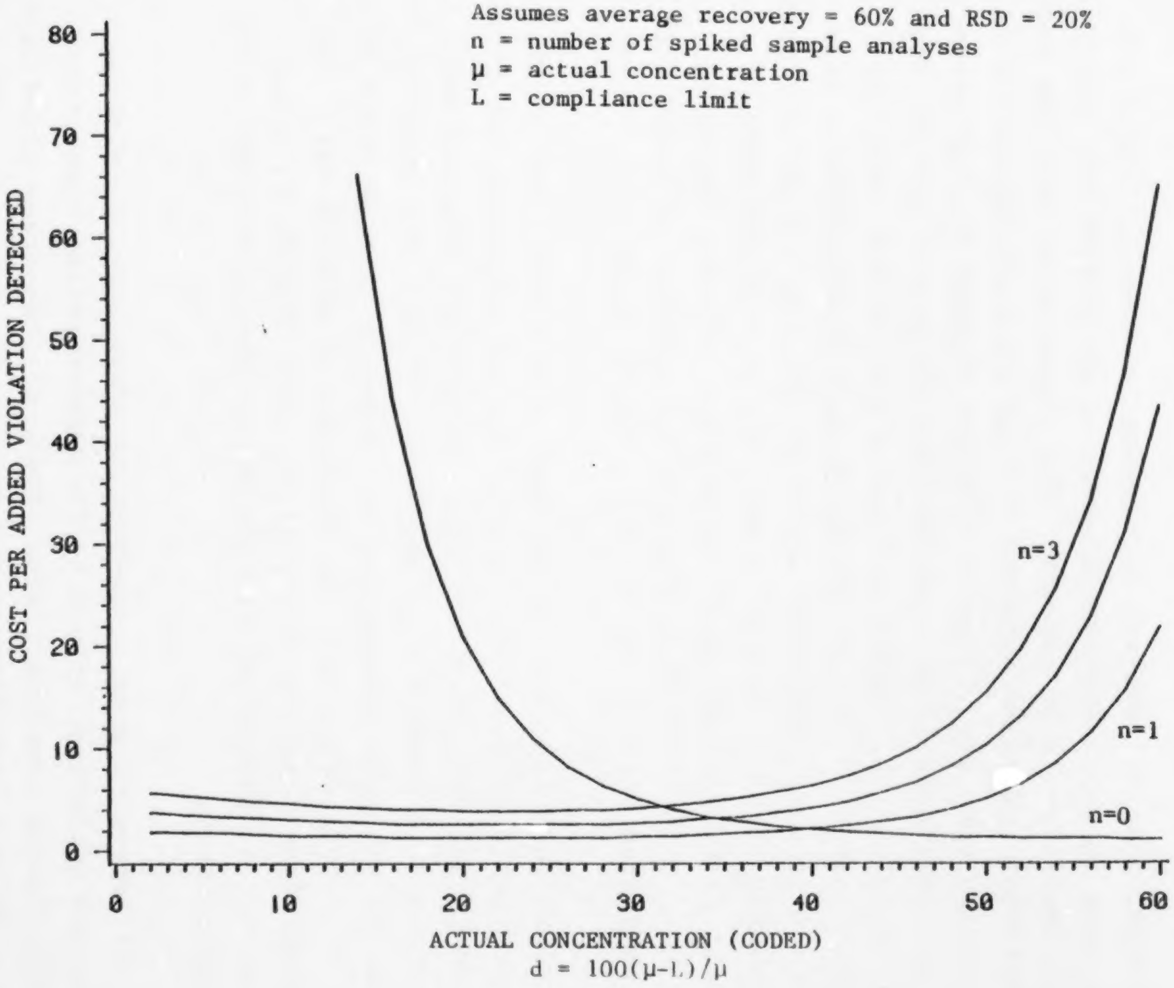


Figure 4-9. Cost-Effectiveness of Bias Correction

limit is 100 ppb. Figure 4-8 shows that the marginal cost of duplicate analyses is about 25 times the cost for a single analysis (d=-0.24 and the curve for n=2 gives MC=25 at this point).

The marginal cost of recovery correction is shown in Figure 4-9 for the case in which average recovery is 60 percent and relative standard deviation is 20 percent. The figure shows that the cost per additional violation detected through recovery correction is less than the cost per violation detected without this QC measure unless noncompliance is so clearcut that analytical quality does not affect detection. Based on Figures 4-8 and 4-9, therefore, it appears that bias correction can be more cost-effective than replication in some regulatory applications. Of course, the cost-effectiveness of bias correction declines as the magnitude of analytical bias declines (e.g., as average recovery approaches 100 percent). Cost-effectiveness for other values of r and RSD can be evaluated through the equations given above.

The marginal cost approach can be used to evaluate other QA/QC activities besides those described if one can determine the sizes of C_0 and C_1 and the impact of QC on the average recovery and relative standard deviation (r and RSD). In general, marginal costs can be put into perspective by comparing them to costs of failing to detect violations (if these can be quantified). The conclusions from marginal cost analyses also depend, as illustrated above, on analytical quality and on concentrations likely to be encountered.

One way of improving compliance-testing efficiency by relating sampling and analytical effort to the concentration encountered is through double sampling plans. These are two-stage plans that allow a compliance decision on the first stage (e.g., based on analysis of a single sample) if results are clearcut (far from the compliance limit). If the first-stage result is near the compliance limit, the second stage (e.g., analysis of another sample) is required and the compliance decision is based on the average of results from both stages. Double sampling plans can be constructed to have the same operating characteristics as single-stage plans (28 - 29). Their advantage is that they require fewer analyses on the average than one-stage plans with comparable operating characteristics, especially when most true concentrations tested are far from the compliance limit.

Evaluating Effectiveness in Comparison Problems

Another common use of analytical data is to compare concentrations resulting from different conditions or sources. Examples of this use are subcategorization of an industry for regulatory purposes (15), treatability studies to compare effectiveness of alternate wastewater treatment systems, and special QC studies to compare recoveries from different instruments in a laboratory. These applications call for statistical tests of whether differences exist between facilities, treatments, or instruments (though estimates of the magnitudes of differences are probably of interest, also).

Making effective comparisons using analytical data depends on proper study planning as much as on analytical quality (30). The quality of comparisons can be affected by the number of analyses, the way samples are allocated to laboratories or analysts for analysis, the order in which samples are analyzed within a laboratory, and the statistical techniques used to evaluate results. The assistance of someone experienced in experimental design, therefore, is helpful in ensuring effective experimental comparisons.

A thorough discussion of measures of effectiveness of comparisons is beyond the scope of this report.* The basic statistical measure of effectiveness for such problems is power: the probability of detecting existing differences of specified size. The evaluation of power for a particular comparison depends on the study plan, the associated statistical model and the values of parameters in the model (e.g., variance components).

The concept of power can be illustrated with the problem of comparing two concentrations; for example, the concentrations of a particular compound in effluents from two manufacturing facilities. The decision of whether the facilities differ in effluent concentration may not be clear-cut because of sampling and analytical variation. Such variation can cause one to conclude that the facilities are different when they are not or are the same when they are different. A statistical test of whether two facilities differ can be designed to limit the probability of wrongfully deciding they are different.** Then for the decision rule so defined, one can compute the probability of detecting differences of given size (i.e., the power of the procedure).

Power curves for common statistical test procedures are readily available (32). They can be used to evaluate the effectiveness of a particular procedure or to choose the most effective study plan to detect specified differences.

To illustrate the use of the power concept in study design, suppose we must determine the number of samples (n) to take from each of two facilities to test for different effluent concentrations. If the total standard deviation of analytical results is σ, then a sample size of approximately

---

*See Davies (31), for example, for more details.

**See Snedecor and Cochran (11), Chapter 4, for details of the two-sample problem.

$$n = 2\sigma^2 (z_\alpha + z_\beta)^2/d^2 \qquad (4.17)$$

is required to detect a difference d with probability (1-β), given that the risk of falsely concluding that the facilities differ is α.* For example, if σ = 10ppb, the desired risk of falsely finding a difference is $\alpha = .05$, and the desired probability of finding a difference of d = 10ppb is $1 - \beta = .95$, then 22 independent samples are needed from each facility.

In general, evaluating the effectiveness of comparisons through power is best done by someone experienced in applied statistics. Success requires ability to identify an appropriate statistical model based on the study design, the nature of the data, and knowledge of the physical and chemical processes involved.

CONCLUDING REMARKS ON QA/QC EFFECTIVENESS

Feedback from users to laboratory quality control programs is invaluable in setting realistic goals for laboratory QA/QC. Experience of users in applying the methods described in the End-Use Quality section will reveal whether shortcomings in analytical quality cause end-use problems, such as necessitating excessive replication to obtain estimates of required precision.

Laboratories, on the other hand, must produce results of consistent, known quality and must communicate quality parameters to users of analytical results to enable them to apply the techniques described in the End-Use Quality section.

---

*See Snedecor and Cochran (1, p.113) for details. z_α and z_β are percentiles of the standard normal distribution, which is tabled in any applied statistics book (e.g., (1)).

REFERENCES

1. Snedecor, G. W. and W. G. Cochran, Statistical Methods, 6th edition, Iowa State University Press, Ames, Iowa, 1967.

2. Owen, D. B., Handbook of Statistical Tables, Addison-Wesley, Reading, Massachusetts, 1962, pp.60-62, 88-99.

3. Anderson, T. W., An Introduction to Multivariate Statistical Analysis, Wiley, New York, 1958, pp.112-115.

4. Bather, J. A., "Control Charts and the Minimization of Costs," Journal of the Royal Statistical Society, B, 25(1), 1963, pp.49-70.

5. Duncan, A. J., "The Economic Design of $\bar{X}$ Charts Used to Maintain Current Control of a Process," Journal of the American Statistical Association, 51(2), 1956, pp.228-242.

6. Hsi, B. P., "Optimization of Quality Control in the Chemical Laboratory," Technometrics, 8(3), 1966, pp.519-534.

7. Currie, L. A., "Limits for Qualitative Detection and Quantitative Determination: Application to Radiochemistry," Analytical Chemistry, 40(3), 1968, pp.586-593.

8. Hubaux, A. and G. Vos, "Decision and Detection Limits for Linear Calibration Curves," Analytical Chemistry, 42(8), 1970, pp.849-855.

9. Kagel, R. O., "Validation and Priority Pollution Analysis," Invited Plenary Address, American Chemical Society National Meeting, Division of Environmental Chemistry, San Francisco, 1980.

10. Environmental Monitoring and Support Laboratory, "Definition and Procedure for the Determination of the Method Detection Limit," U.S. EPA, Office of Research and Development, Cincinnati, Ohio, January, 1981.

11. Glaser, J. A., et al, "Trace Analyses for Wastewaters." Environmental Science and Technology, 15(12), 1981, pp. 1426-1435.

12. Long, G. L. and J. D. Winefordner, "Limit of Detection - A Closer Look at the IUPAC Definition," Analytical Chemistry, 55(7), 1983, pp. 712A-724A.

13. Freund, J. E., *Modern Elementary Statistics*, 5th edition, Prentice-Hall, Englewood Cliffs, New Jersey, 1979.

14. MacDougall, D., et al., "Guidelines for Data Acquisition and Data Quality Evaluation in Environmental Chemistry," *Analytical Chemistry*, 52(14), 1980, pp.2242-2249.

15. Horwitz, W., L. R. Kamps and K. W. Boyer, "Quality Assurance in the Analysis of Foods for Trace Constituents," *Journal of the Association of Official Analytical Chemists*, 63(6), 1980, pp.1344-1354.

16. Cochran, W. G., *Sampling Techniques*, 2nd edition, Wiley, New York, 1963.

17. Bennett, C. A. and N. L. Franklin, *Statistical Analysis in Chemistry and the Chemical Industry*, Wiley, New York, 1954.

18. Environmental Protecton Agency, "Timber Products Point Source Category," *Federal Register*, 46(16), January 26, 1981, p.8263.

19. Huibregtse, K. R. and J. H. Moser, *Handbook for Sampling and Sample Preservation of Water and Wastewater*, Envirex, Inc., EPA-600/4-76-049, U.S. EPA, Office of Research and Development, Environmental Monitoring and Support Laboratory, Cincinnati, Ohio, 1976.

20. Nolan, T. W., R. S. Elder and M. L. Hereth, *NSPS for SO Emissions from Industrial Boilers - Statistical Issues and Analysis of FGD Data*, Radian Corporation, EPA Contract No. 68-02-3058, Office of Air Quality Planning and Standards, Research Triangle Park, NC, 1981.

21. Duncan, A. J., "Bulk Sampling: Problems and Lines of Attack," *Technometrics*, 4, 1962, pp.319-344.

22. Rhode, C. A., "Composite Sampling," *Biometrics*, 32, 1976, pp.273-282.

23. Schaeffer, D. J., H. W. Kerster and K. G. Janardan, "Grab Versus Composite Sampling: A Primer for the Manager and Engineer," *Environmental Management*, 4(2), 1980, pp.157-163.

24. Elder, R. S., W. O. Thompson and R. H. Myers, "Properties of Composite Sampling Procedures," *Technometrics*, 22(2), 1980, pp.179-186.

25. Hill, I. D., "The Economic Incentive Provided by Sampling Plans," Applied Statistics, 9, 1960, pp.69-81.

26. Rice, J. K., "Analytical Issues in Compliance Monitoring," Environmental Science and Technology 14(12), 1980, pp. 1455-1457.

27. Duncan, A. J., Quality Control and Industrial Statistics, 4th edition, Richard D. Irwin, Inc., Homewood, Illinois, 1974.

28. Bartlett, R. P. and L. P. Provost, "Tolerances in Standards and Specifications," Quality Progress, 6(12), 1973, pp. 14-19.

29. Neuhauser, D. and A. M. Lewicki, "What Do We Gain from the Sixth Stool Guaiac?", New England Journal of Medicine, 293, 1975, pp.226-228.

30. Elder, R. S., "Double Sampling for Lot Average," Technometrics, 16(3), 1974, pp.435-439.

31. Hald, A., "Optimum Double Sampling Tests of Given Strength I. The Normal Distribution," Journal of the American Statistical Association, 70(2), 1975, pp.451-456.

32. Health Effects Research Laboratory, "Development of Quality Assurance Plans for Research Tasks," EPA-600/1-78-012, U.S. EPA, Office of Research and Development, Research Triangle Park, NC, 1978.

33. Davies, O. L., The Design and Analysis of Industrial Experiments, 2nd edition, Hafner Publishing Co., New York, 1956.

34. Beyer, W. H. (editor), Handbook of Tables for Probability and Statistics, 2nd edition, Chemical Rubber Co., Cleveland, Ohio, 1968.

SECTION 5

CHOOSING COST-EFFECTIVE QA/QC PROGRAMS

Designing an effective quality control program for an official analytical method is complicated by the following factors:

- The method will be applied in many laboratories, each with potentially different costs and quality problems
- There may be many different uses and users of analytical results, each with possibly different quality needs
- Users may be unable to specify quality needs (at least until experience is gained in the use of a method)
- The method may measure many parameters on each sample; quality control needs for different parameters may conflict.

In such a setting, a single QA/QC program cannot be cost-effective for every application. One reasonable way to avoid requiring superfluous efforts is to establish a two-tier program on the following basis:

- Minimal - QC steps needed regardless of use
- Additional - QC steps tailored to end-use needs

The cost-effectiveness of QA/QC programs can be improved further by basing levels of QC effort in both the minimal and additional phases on reasonable QC targets (as discussed in Section 4). The development of a minimal program is discussed in the Minimal QA/QC Programs part of Section 5; the selection of additional QA/QC procedures is discussed in the Additional QA/QC Efforts part of Section 5. The USEPA 600 series methods are used to illustrate the principles discussed in these sections.

MINIMAL QA/QC PROGRAMS

Establishing a minimal QA/QC program requires:

- Identifying possible uses of data
- Identifying quality needs common to all uses
- Selecting QA/QC activities to satisfy the common needs
- Picking appropriate levels of effort for each selected activity

The results of this identification process will differ among analytical methods, so a detailed quantitative prescription cannot be given. However, qualitative guidance is provided, with examples using USEPA's 600 series methods for wastewater analysis.

The needs of a minimal QA/QC program include:

- Organization
- Appraisal
- Process control
- Interlaboratory compatibility

"The most important factor in setting up quality control is the establishment of the organization to do the job (1)." Sources of information on quality control organization were listed in Section 3.

The development and implementation of a laboratory quality control program must begin with a review of the goals and philosophies of the laboratory's management. Quality objectives can be quite different, depending on management's objectives for a QC program. Laboratory QC goals will be related to the organization's position and reputation. QC goals could be oriented to:

- Minimum cost for quality control,
- Generation of laboratory data of known quality,
- Generation of laboratory data at an established quality level,
- Maintenance of schedules (productivity), and
- Matching quality goals to end-use needs.

The quality objectives of management must be stated prior to implementing a specific quality control program. In addition to stating objectives, management's quality goals will be defined by QA organization, QC budgets, and other restrictions on potential QA/QC activities.

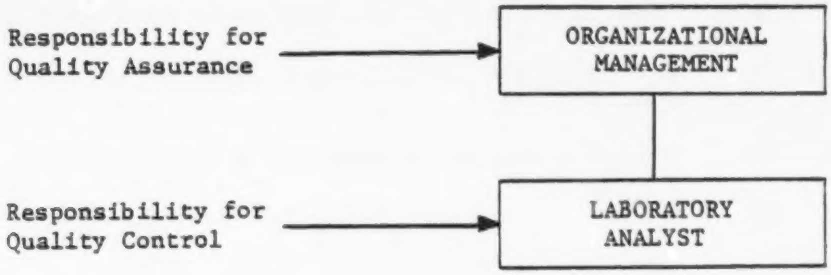
The concepts of total quality control (1) involve all relevant personnel in QA/QC activities from laboratory managers and analysts to users of analytical data. The concept that everyone in the organization affects quality is good, but clear accountability and responsibility for quality control must also be delineated. Personnel performing the quality functions should have sufficient authority and organizational freedom to identify quality problems and initiate solutions.

Laboratories which use the 600 series methods for self-monitoring or regulation will include:

- Industrial laboratories,
- Contract laboratories, and
- Government laboratories.

Each of these types of laboratories ranges in size from a single analyst/single equipment laboratory to a large laboratory with many analysts and types of equipment. Figure 5-1 depicts possible organizational sructures for these extremes. For the very

1. One Instrument/One Analyst Laboratory:


Responsibility for Quality Assurance
ORGANIZATIONAL MANAGEMENT
Responsibility for Quality Control
LABORATORY ANALYST

2. Large Laboratory:

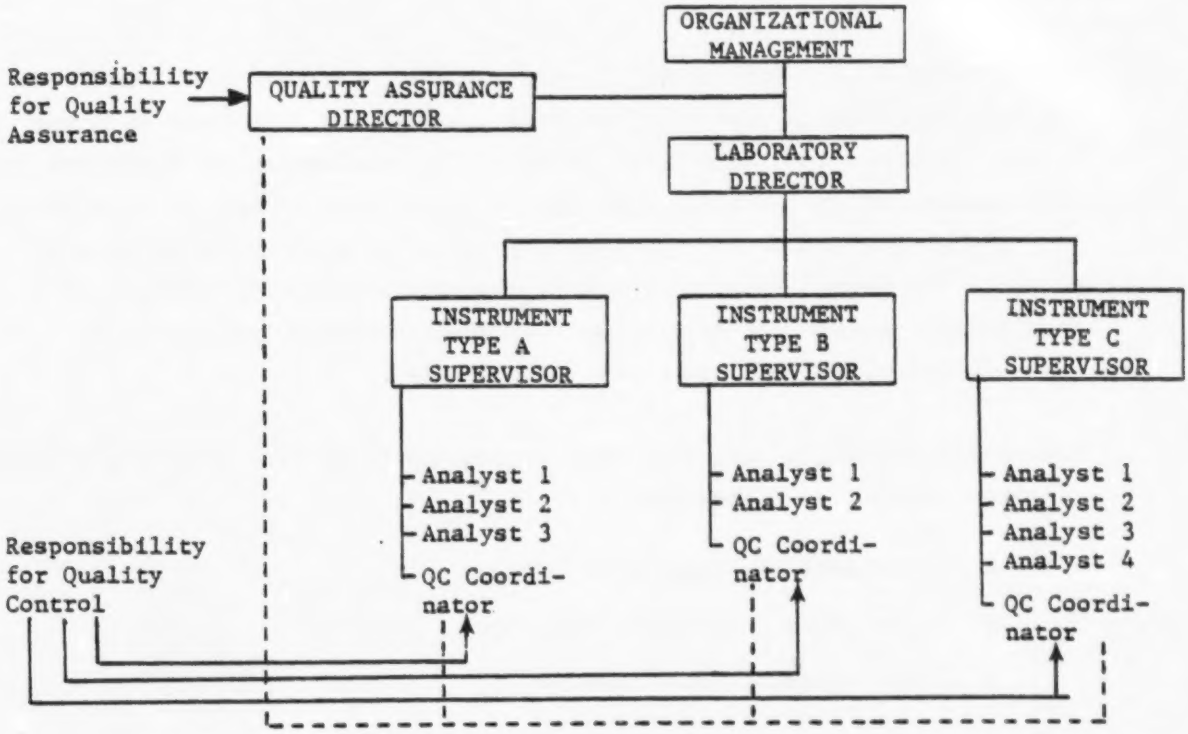

ORGANIZATIONAL MANAGEMENT
Responsibility for Quality Assurance
QUALITY ASSURANCE DIRECTOR
LABORATORY DIRECTOR
INSTRUMENT TYPE A SUPERVISOR
INSTRUMENT TYPE B SUPERVISOR
INSTRUMENT TYPE C SUPERVISOR
Analyst 1
Analyst 2
Analyst 3
QC Coordinator
Analyst 1
Analyst 2
QC Coordinator
Analyst 1
Analyst 2
Analyst 3
Analyst 4
QC Coordinator
Responsibility for Quality Control

Figure 5-1. Quality Control Organizations for Laboratories

small laboratory, the single analyst is responsible for implementing, analyzing, and reporting all quality control, while the organizational management assumes quality assurance responsibility. For the large laboratory, a Quality Assurance Director reporting to the organizational management is responsible for QA activities. Quality Control Coordinators are designated in each laboratory to implement the specific QC activities in each area. These QC coordinators report QC results and summaries to the QA Director, as well as to the supervisor in their specific area.

Other QA/QC organizations can be developed which fit the specific organizational structure of the laboratory. The key requirement is that QA/QC personnel have the responsibility and organizational freedom to effectively implement the QC program.

Each laboratory using the 600 series methods should have a Quality Control Manual. The use of the 600 series methods can be incorporated in the laboratory's general QC manual or a specific manual can be prepared for these applications. The manual should document all aspects of the laboratory QA/QC, including each of the following areas:

1. QA/QC policies and objectives
2. QA/QC organization and personnel
3. Sampling and sample custody procedures
4. Analytical methods and method validation
5. Calibration procedures
6. Quality control tests and frequency
7. Data handling, validation, and reporting
8. QC reporting, review, and corrective action procedures
9. QA auditing procedures

The QC manual should be kept up to date as personnel, method, and procedural changes occur.

Appraisal of quality is necessary to document that satisfactory results are being produced, to detect quality problems and to provide information needed by users for planning and evaluation. The most important measures of analytical quality under laboratory control are bias and precision. Other quality measures of interest are frequencies of false positives or negatives (detection problems). Appraisal alone is of little value, however, because without process control there is no fixed quality to document. Process control is needed to ensure that quality estimates obtained at one time reflect the quality produced at other times (which is essential for planning purposes).

The average recovery demonstrated in method validation should be the ultimate laboratory target (i.e., zero bias relative to the average for the method). However, achieving a constant, documented average recovery (even if different from the method average) is an important first step in analytical QC, because within-laboratory control is a prerequisite to controlling between-laboratory differences. In addition, for applications such as NPDES compliance monitoring, bias corrections can be made effectively if the amount of bias is well documented (which generally will be possible only if bias is constant over time).

The appropriate precision target depends on circumstances of analysis and use. Within-day, between-day and longer run variance components in a laboratory are the basis for precision targets. Benchmark estimates of these parameters should be obtained in method validation studies. In NPDES monitoring, the importance of the magnitude of analytical precision declines the farther the true concentration is from the compliance limit. However, maintaining a constant level of precision is important in every case.

Compatibility of results produced in different laboratories is necessary for comparisons of results over space or time to be meaningful (2, 3). Whether special steps are necessary to achieve interlaboratory compatibility depends on the test method. Validation studies for most analytical methods indicate that interlaboratory differences are an important source of error.

When data is generated by more than one laboratory for the same purpose and relative biases between laboratories cannot be eliminated, it sometimes is possible to allocate analyses to laboratories so that differences of interest can be separated from between-laboratory differences. However, within-laboratory control of bias is important in this case too.

Interlaboratory QC customarily consists of periodic analyses of round-robin samples provided by a coordinating laboratory to check for relative bias between laboratories. Quarterly or annual samples are sometimes specified. One potential problem with the round-robin approach in the 600 method case is the vast number of different matrices with which laboratories must deal. A possibly more practical approach to controlling interlaboratory bias is a program to insure that laboratories use equivalent standards, combined with use of methods in which the equivalence of any options is conclusively demonstrated.

Any quality control program must be a combination of preventive measures that keep quality problems from occurring and of QC tests that provide feedback on the quality of analytical results being produced. It generally is more cost-effective to concentrate efforts on prevention rather than appraisal, but some testing is always needed.

The best means of achieving QA/QC needs depends on the tools available for a method. For example, standard reference materials, one widely recommended appraisal tool, are not available for many environmental applications. The best choice also depends on resources available and the kinds of quality problems that prevail.

The best approach to choosing the types of QC tests to run routinely from those possible is to select the minimum number needed to detect problems of the types and sizes that are important to detect. This is accomplished by selecting those tests that are most comprehensive in the types of problems their results will reflect. When these tests indicate the presence of a quality problem, then one can run supplemental tests if necessary to identify the source of the problem.

To illustrate this approach, the types of tests recommended for a minimal QC program for the 600 series methods would include:

- Response factor stability - reflects changes in instruments or standards
- Spiked sample recoveries - reflects recovery bias (including contamination) and long-run recovery variation
- Duplicate analyses - reflects short-run recovery variation (precision)

Field spiking into reagent water would be the best procedure, if practicable, since field spikes are exposed to more potential problems than laboratory spikes. Use of reagent water avoids background problems, so is preferable (in the absence of matrix effects). Duplicate spiked samples are preferable to duplicate environmental samples (again, if absence of matrix effects permits) because of the difficulty of ensuring that duplicate environmental samples contain the analytes of interest. Use of spike recoveries of surrogate compounds is another type of QC test that has potential value for routine use, provided suitable surrogates are available for a method.

The minimal approach just described can be supplemented with other types of tests, when necessary, to discover the source of a problem. For example, the cause of bias uncovered by a spiked sample recovery might be identifiable by analyzing a laboratory blank to check for contamination.

The approach just described contrasts with the comprehensive approach in which periodic tests are done to check for every different kind of potential problem; e.g., reagent waer and matrix spikes, field and method blanks, standard checks, response factor tests, etc. (4). The drawback of the comprehensive approach is that it requires extensive testing in all laboratories, whether they experience quality problems or not. Thus, it can require a great deal of wasted QC-testing effort in laboratories that take effective action to prevent quality problems from occurring. Careful use of a very few different kinds of QC tests can provide sufficient quality documentation and general quality appraisal in any laboratory, provided additional tests are performed when needed for diagnostic purposes.

ADDITIONAL QA/QC EFFORTS

Elements of a minimal program applicable to all laboratories were described in the previous section. However, "quality control procedures will ordinarily need to be custom made for each situation and, perhaps, for each laboratory (5)." A detailed discussion of how to tailor QA/QC programs for specific uses was given in Sections 3 and 4. (For other sources of information on QA project plans, experimental design and general quality control, see references (6) to (12)). A general description and two examples are given below to illustrate the design process.

The Decision Process

Figure 5-2 is a flowchart outlining steps involved in tailoring a quality control program for a particular use. The central fact is that specific, quantitative objectives can be identified that permit the use of the statistical design methods described in Section 4. The initial application of these methods may indicate the need for an overly costly program; if so, quality needs will have to be re-examined to identify changes that permit a feasible program. Successful program design requires interaction between laboratories and users of results.

The selection of the analytical method will be the limiting factor for many types of chemical analysis. Sometimes ony one accepted method is available. The 600 series methods provide alternative methods for analyzing a specific pollutant. The basic choice is between one or more of the gas-liquid or high performance liquid chromatography methods or one or more of the mass spectrometer methods. Some of the methods include optional procedures, apparatus, and materials. Modifications of the methods, beyond those expressly permitted in the written method procedures, are considered as "major" modifications. Any such modifications must be approved as an alternate test procedure.

The choice of a GC versus a GC/MS method for each pollutant offers the individual laboratory flexibility in selection of methods. Compared to the GC/MS methods, the GC methods are simple, require inexpensive equipment, and do not require sophisticated operators. The GC/MS methods generally require less sample clean-up and preparation. GC methods are most advantageous when the laboratory has had prior experience with the sample matrix, the matrix is relatively clean, and only a few of the priority pollutants are expected to be present in the sample.

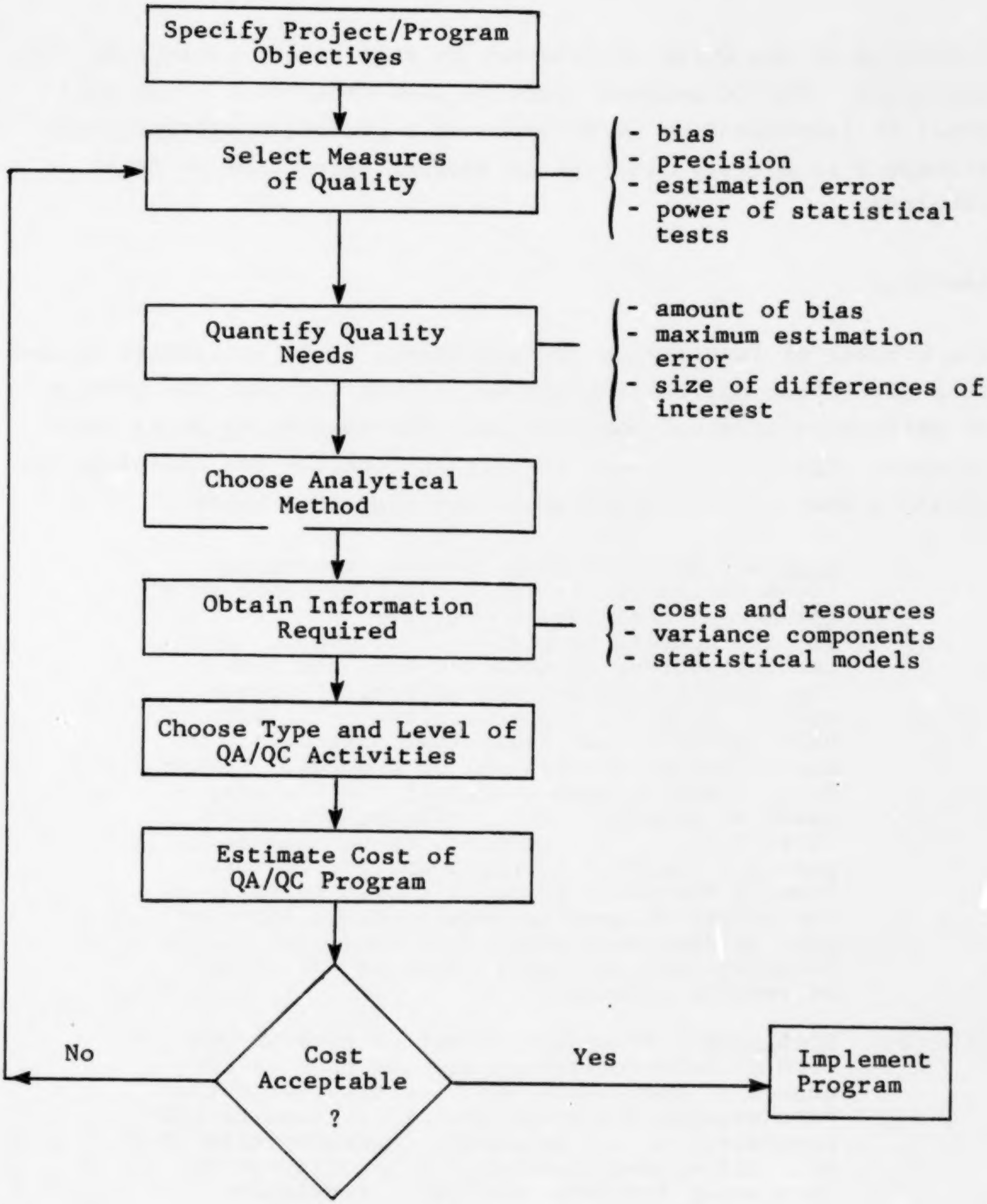

Specify Project/Program Objectives
Select Measures of Quality
- bias
- precision
- estimation error
- power of statistical tests
Quantify Quality Needs
- amount of bias
- maximum estimation error
- size of differences of interest
Choose Analytical Method
Obtain Information Required
- costs and resources
- variance components
- statistical models
Choose Type and Level of QA/QC Activities
Estimate Cost of QA/QC Program
Cost Acceptable ?
No
Yes
Implement Program

Figure 5-2. Steps Involved in Tailoring a Quality Control Program for a Particular Use

Combining GC and GC/MS methods may be cost-effective in some laboratories. The GC methods could be used routinely. Samples could be reanalyzed by GC/MS whenever a pollutant concentration exceeds a regulatory limit or an unusual pattern of GC peaks is observed.

Examples

The process of identifying the additional QA/QC activities needed in a particular application can be illustrated with the problem of performing chemical analyses for self-monitoring under NPDES permits. The quality needs in this application and the QA/QC activities that satisfy these needs are discussed below.

1. Bias - Underestimating effluent concentrations increases the chance of compliance at a given concentration, so negative bias is the primary concern in permit enforcement. Potential sources of negative bias are sampling, sample-handling and analytical procedures. Sampling procedures are dealt with in NPDES permits, and analytical bias should be controlled by the minimal QC program. Therefore, handling loss apparently is the only cause of negative bias requiring additional QC activity. This problem can be checked by periodic analysis of field-spiked samples. Testing frequency probably should be based on the number of samples between tests (the impact of incorrect compliance decisions due to handling loss is proportional to the number of samples affected).

2. Precision - Precision is not of as much concern in self-monitoring as bias because the chance of compliance when the true concentration exceeds the compliance limit remains low regardless of the precision (assuming bias is negligible; see Evaluating Effectiveness in Regulatory Problems section). Precision-control measures in the minimal program should be adequate in this application.

3. Contamination - The effect of contamination is to increase the apparent concentrations in monitoring samples, so contamination has an adverse impact on permittees. There are two ways to deal with contamination. One option is to test blanks and use the blank readings to correct sample readings for contamination. This approach increases the variation in monitoring test results when the contamination level is high enough to affect compliance decisions. Therefore, it can decrease the permittee's chance of compliance at some acceptable concentrations and decrease the chance of detecting noncompliance at higher concentrations. The other option is to eliminate the causes of contamination. The permittee probably should be allowed to choose a strategy for handling contamination, since requiring the analysis of blanks is not cost-effective for laboratories without contamination problems. Cost-effectiveness also will be affected by laboratory size and frequency of blank analyses.

In general, the decision to require a particular QA/QC activity should depend on how the quality problem addressed by that activity affects the permittee. If a problem such as contamination makes compliance more difficult, it probably is not necessary to require QC activities to control the problem. Pointing out options, but allowing permittees to choose a solution, is the cost-effective approach. If a problem such as handling loss makes compliance easier, however, QC steps aimed at controlling the problem should be required. By this reasoning, analysis of field-spiked samples is probably the only additional QC activity warranted in self-monitoring for NPDES permits.

A second example of identifying additional QA/QC needs involves the use of analytical results from treated wastewater samples to set effluent limitation guidelines (see (13) for an illustration). The objective is to set limits that can be achieved a high percentage of the time by facilities in a particular industrial category using appropriate treatment technology. This requires making reasonable allowance for between-facility, process, sampling and analytical variation. This was done in reference (14) by estimating the 99th percentile of effluent concentrations for several facilities. An appropriate measure of end-use quality in this case would be the error in estimating the 99th percentile from sample data subject itself (possibly) to both random and systematic errors. Since enforcement would be based on the method used to set the guidelines, method bias is not a problem. Biases due to sample loss or contamination, however, could result in unrealistic limits, so these sources of bias should be checked (e.g., through field blanks and field spikes). Another concern is to obtain samples that adequately reflect the variation characteristic of each facility. The sampling strategy that can be followed depends on the cost of obtaining and analyzing samples (the high cost of analyzing organic priority pollutants typically limits sampling to three consecutive daily samples (15); several years' daily samples may be available for other parameters (14)). A third concern is the possibility that interlaboratory variation could become confused with other sources of variation, such as facilities or samples. This problem can be avoided by effective assignment of samples to laboratories. A final consideration is the choice of appropriate statistical methods for data analysis. For example multivariate analysis can be used instead of independent analysis of each compound (see (14) for another example). In summary, the additional quality control tools apparently needed in the effluent guidelines problem are field blanks and spikes, study planning, and appropriate statistical estimation methods.

Other examples of selecting QA/QC efforts are given in References (16 - 18).

REFERENCES

1. Fiegenbaum, A. V., Total Quality Control, McGraw-Hill, New York, 1961.

2. Uriano, G. A. and C. C. Gravatt, "The Role of Reference Materials and Reference Methods in Chemical Analysis," CRC Critical Reviews in Analytical Chemistry, 6(4), 1977, pp.361-411.

3. Environmental Monitoring and Support Laboratory, Handbook for Analytical Quality Control in Water and Wastewater Laboratories, EPA-600/4-79-019, U.S. EPA, Office of Research and Development, Cincinnati, 1979, p.1-2.

4. Versar, Inc., "Quality Assurance for Laboratory Analysis of 129 Priority Pollutants," (Interim Report), EPA Contract No.68-01-5948, U.S. EPA, Office of Water Planning and Standards, Washington, D.C., February, 1980.

5. Taylor, J. K., "Validation of Environmental Data by Intercalibration and Laboratory Quality Control Programs," Presented before the American Chemical Society, Division of Environmental Chemistry, Los Angeles, CA, 1974.

6. Environmental Monitoring and Support Laboratory, Handbook for Analytical Quality Control in Water and Wastewater Laboratories, EPA-600/4-79-019, U.S. EPA, Office of Research and Development, Cincinnati, 1979.

7. Quality Assurance Management Staff, "Interim Guidelines and Specifications for Preparing Quality Assurance Project Plans," QAMS-005/80, U.S. EPA, Office of Research and Development, Office of Monitoring Systems and Technical Support, Washington, D.C., 1980.

8. Health Effects Research Laboratory, "Development of Quality Assurance Plans for Research Tasks," U.S. EPA, Office of Research and Development, Research Triangle Park, NC, 1978.

9. Natrella, M. G., Experimental Statistics, NBS Handbook 91, U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., 1966.

10. Davies, O. L., The Design and Analysis of Industrial Experiments, 2nd edition, Hafner Publishing Co., New York, 1956.

11. Duncan, A. J., Quality Control and Industrial Statistics, 4th edition, Richard D. Irwin, Inc., Homewood, IL, 1974.

12. Fiegenbaum, A. V., *Total Quality Control*, McGraw-Hill, New York, 1961.

13. Juran, J. M. and F. M. Gryna, *Quality Planning and Analysis*, McGraw-Hill, New York, 1970.

14. Environmental Protection Agency, "Timber Products Point Source Category," *Federal Register*, 46(16), January 26, 1981, pp.8260-8295.

15. Holtzclaw, P. W. and M. D. Neptune, "Approach to Quality Assurance/Quality Control in the Organic Chemicals Monitoring Program," *Journal of Environmental Science and Health*, A15(5), 1980, pp.525-543.

16. Taylor, J. K., "Quality Assurance of Chemical Measurements," *Analytical Chemistry* 53 (14), 1981, pp. 1588A-1596A.

17. Dux, J. P. "Quality Assurance in the Analytical Laboratory," *American Laboratory* 63, 1983, pp. 54-63.

18. Aldenhoff, G. A. and L. A. Ernest, "A Quality Assurance Program at a Municipal Wastewater Treatment Plant Laboratory," *Journal of Water Pollution Control Federation* 55 (9), 1983, pp. 1132-1137.

APPENDIX A

SKIP-LOT PROCEDURES

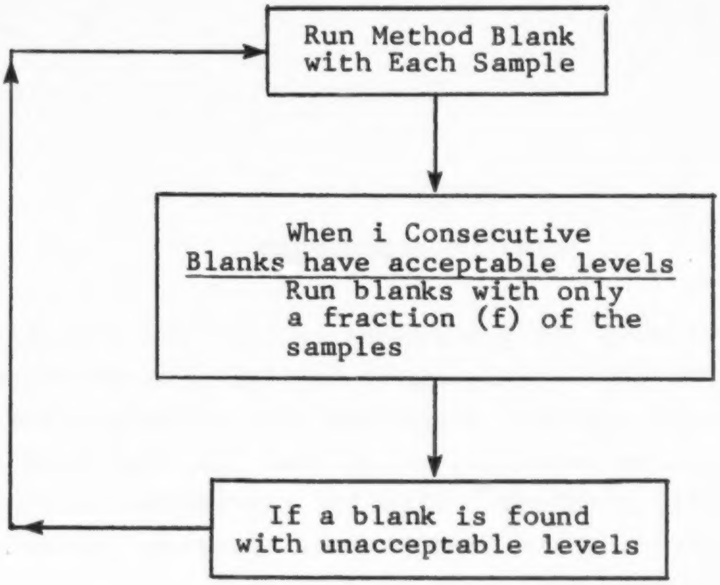
The basic idea of skip-lot procedures is that the amount of appraisal effort required in a quality control program depends on the quality being produced. A process that produces consistently high quality requires less monitoring than one that frequently experiences quality problems. Skip-lot procedures provide objective rules for deciding the frequency of appraisal needed. They are applicable to any continuous production process in which the items produced are expected to be of similar quality, and quality is not deliberately changed depending on the level of appraisal.

The basic procedure was developed by Dodge (1).* It is illustrated in Figure A-1 in terms of analyzing blanks to check for contamination. The procedure switches between analyzing one blank with every sample and analyzing one blank with every f samples; the current level depends on whether recent performance has been acceptable. The procedure is described by the parameters i and f (defined in the figure).

One statistical property of skip-lot procedures, which is useful because it reflects their economic impact, is the average fraction inspected (in the case of blanks, this is the average percentage of samples with which blanks are analyzed). It can be shown that when quality is unacceptable, the skip-lot procedure requires a blank to be analyzed with every sample (or group of

---

*The original procedure as applied to individual units was called a "continuous sampling plan" by Dodge.


Run Method Blank with Each Sample
When i Consecutive Blanks have acceptable levels
Run blanks with only a fraction (f) of the samples
If a blank is found with unacceptable levels

Alternative Skip-Lot Plans

AOQL* = 10%		AOQL = 5%	
i	f	i	f
6	1/4	11	1/4
7	1/5	14	1/5
8	1/7	18	1/7
11	1/10	21	1/10

*AOQL = maximum average percentage of samples with contamination undetected due to skipping.

Figure A-1. Skip-Lot Sampling Plan Applied to the Analysis of Method Blanks samples). When quality is very good, on the other hand, a blank will be required with only one in f samples (or sample groups). Intermediate quality levels result in intermediate average fractions inspected.* This dependence of appraisal costs on production quality provides an economic incentive to produce good quality.

Another statistical property of skip-lot procedures is the average outgoing quality (AOQ) as a function of quality produced. In the case of blanks, the AOQ is the expected average percent of samples with unacceptable contamination that are undetected due to skipping (testing at rate f). This property also depends on the level of quality being produced. However, it can be shown mathematically that the AOQ for a given skip-lot plan has a maximum value, called the AOQL. The AOQL for blanks would be the maximum average percentage of samples with unacceptable contamination undetected due to skipping. Skip-lot plans usually are chosen based on AOQL's. Some plans with AOQL's of 5 and 10 percent are shown in the figure.

Dodge and Perry (2, 3) developed skip-lot procedures to apply to collections of units which are tested based on a subsample of the units. This procedure could be applied to batches of samples; for example, a blank could be run with each batch until sufficient acceptable results were obtained, then one in f batches could be tested. This approach would be appropriate when quality problems tend to affect all samples in a batch similarly (e.g., field contamination).

---

*It is possible to derive a mathematical expression for the average fraction inspected in terms of i, f and the probability of detecting a problem in an individual test (1).

A further generalization of skip-lot procedures allows more than one level of skipping inspection (4). This approach can permit further savings, but is more difficult to administer (especially when there are several quality parameters of interest; e.g., several potential contaminants to check).

In conclusion, skip-lot procedures can be developed for monitoring any monitorable aspect of analytical quality. They result in the greatest savings when high-quality production is common. They provide an incentive to produce high quality as long as quality requirements are set at achievable levels.

The skip-lot philosophy is incorporated in the Environmental Protection Agency's Test Procedures for the analysis of pollutants in wastewater (5). The frequency of spiked sample and QC check sample analysis is dependent on quality performance. After demonstrating the ability to perform acceptable analysis, the "start-up test", QC checks are reduced to ten percent of the samples analyzed. Further reduction is possible if all test criteria are met.

REFERENCES

1. Dodge, H. F., "A Sampling Inspection Plan for Continuous Production," Annals of Mathematical Statistics, 14(3), 1943, pp. 264-279.

2. Dodge, H. F., "Skip-Lot Sampling Plan," Industrial Quality Control, 11(5), 1955, pp.3-5.

3. Perry, R. L., "Skip-Lot Sampling Plans," Journal of Quality Technology, 5(3), 1973, pp.123-130.

4. Perry, R. L., "Two-Level Skip-Lot Sampling Plans - Operating Characteristic Properties," Journal of Quality Technology, 5(4), 1973, pp.160-166.

5. Evironmental Protection Agency, "Guidelines Establishing Test Procedures for the Analysis of Pollutants Under the Clean Water Act," Federal Register, 49(209), October 4, 1984, pp.43234-43406.

APPENDIX B

DESIGN AND ANALYSIS OF SPIKE-RECOVERY STUDIES

In spiked (fortified) sample studies, known amounts of a compound or compounds of interest are added to aliquots of a sample, and the percentage of analyte recovered by a test method is used to evaluate the performance of that method. The Environmental Protection Agency (EPA), for example, uses spiking studies in method development (e.g., (1)) and has proposed the use of spiked samples in quality control programs under National Pollutant Discharge Elimination System (NPDES) permits (2). Thus the proper conduct and interpretation of spiking programs is critical to the development and implementation of the analytical methods upon which important environmental programs are based.

Spiking is particularly useful in wastewater analyses because the variety of sample matrices and the number of analytes of interest in each sample make realistic standard reference materials difficult to produce. Spiking permits flexibility in the choice of sample matrix and in the combinations and levels of analytes that can be evaluated. The usefulness of spiked-sample analyses is not limited to wastewater or environmental samples, however, and proper interpretation of data from such analyses (percent recovery data) is important whatever the application. Analytical results from these studies usually are evaluated in terms of percent recoveries of the spiked material. The possibility of nonzero background levels in spiked samples raises the following questions:

- How should percent recovery be defined and calculated?
- What are the best spike levels in relation to background concentrations?

In this Appendix, statistical properties of percent recovery data, when analytical bias and precision are proportional to sample concentration, are described. The impact of the presence of the analyte of interest in the unspiked sample (i.e., nonzero background concentration) is examined and some of the potential pitfalls in the interpretation of percent recovery data in method development and quality control applications are discussed.

DEFINITION OF PROBLEM

It is commonly found that in the region of applicability of methods of trace analysis, both precision and bias are proportional to concentration. That is, analytical results have mean and variance

$$\begin{aligned} E(X) &= pB \\ V(X) &= (pBC_A)^2, \end{aligned} \qquad (B.1)$$

where B is the true sample concentration, 100p is the percent recovery of the method, and $100C_A$ is the analytical coefficient of variation (or RSD). An important function of within-lab QC is to keep p and C_A constant; ideally, p=1 and C_A is small.

Percent recovery often is controlled through the analysis of spiked or fortified samples by comparing measured concentrations to spike levels. One must understand the statistical properties of percent recovery data to use it effectively. Statistical properties of such data depend on the following factors:

- the value of initial sample concentration (background concentration),
- the method of determining spike level (fixed or proportional to background), and
- the magnitude of spike employed.

These factors should be taken into account in interpreting percent recovery data but often are not.

A number of definitions of percent recovery have been used in QA/QC programs. The following three definitions are considered here (1):

Definition 1: $R_1 = 100\ Y/T$ (zero background)

Definition 2: $R_2 = 100\ (Y-X)/T$ (nonzero background)

Definition 3: $R_3 = 100\ (Y-X)/hX$ (nonzero background)

where

X = measured background concentration

Y = measured spiked-sample concentration

T = increase in concentration due to spiking when spike level is fixed

hX = increase in concentration due to spiking when spike level is h times measured background.

The first definition applies (for example) to spiked reagent water samples for which background concentration is zero (unless there is contamination). This definition is often applicable in collaborative studies that determine the p and C values achievable with an analytical method. The second and third definitions apply (for example) to wastewater samples with positive background concentrations. These definitions are applicable in QC of routine sample analyses to control percent recovery in natural matrices.

ZERO BACKGROUND - DEFINITION 1

If the original sample concentration is zero and the spike level is T, the assumed properties of the measurement process (B.1) can be used to show that

$$\begin{aligned} E(R_1) &= 100p \\ CV(R_1) &= 100C_A \end{aligned} \qquad (B.2)$$

Thus recovery data of this type accurately reflects the assumed properties. This is the only case in which this will prove to be true.

Definition 1 can give misleading results if presumably clean samples are contaminated. If samples are contaminated with amount B, then the true concentration of spiked samples is B+T and

$$\begin{aligned} E(R_1) &= 100p(1+B/T) \\ &= 100p + (100pB)T^{-1} \end{aligned} \qquad (B.3)$$

In this case, R_1 is a biased estimator of 100p; the bias decreases with increasing T. Using an average of such results to set control limits gives unrealistic limits at low T values. If results are available for two or more spike levels (T values), one can regress R_1 on T^{-1}; the intercept is an unbiased estimator of 100p. This approach is useful for some organic compounds (e.g., phthalates) that are very difficult to keep from contaminating laboratory environments.

Contamination changes the variance of R_1 to

$$V(R_2) = (100pC_A)^2(1+B/T)^2 \qquad (B.4)$$

Note that variation in R_1 is large at low spike levels, but approaches the assumed value as T increases. The coefficient of variation equals the assumed value for any T, however.

The contamination level B was assumed to be constant in (B.3) and (B.4). If B varies randomly from sample to sample, the bias of R_1 is a function of the mean contamination level. The variance of R_1 reflects variation in contamination as well as analytical variation. The coefficient of variation no longer has the correct value.

NONZERO BACKGROUND - DEFINITION 2

If background concentration B is constant, using (B.1)

$$E(R_2) = 100p$$

and $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (B.5)

$$V(R_2) = (100pC_A)[(B/T)^2+(1+B/T)^2]$$

Thus R_2 has the correct mean value regardless of B or T, but $V(R_2)$ depends on the ratio of B and T. As T/B increases, $V(R_2)$ approaches the correct value.

The consequence of this result is easily seen through some examples. Table B-1 shows the impact of T/B on Var (R_2) and the expected range in recoveries for three cases $(m = n = 1)$. The expected range in recovery is based on a 95 percent tolerance interval for a normal distribution:

$$100\ p \pm 1.96 \sqrt{Var\ (R_2)}$$

TABLE B-1. IMPACT OF SPIKE TO BACKGROUND RATES ON VARIABILITY OF PRECENT RECOVERIES (Definition 2)

Spike to Background Ratio (T/B)	Var (R_2)	Expected Range in Percent Recoveries*		
		($p = 1.0$ $C_A = 0.1$)	($p = 1.0$ $C_A = .2$)	($p = .5$ $C_A = .2$)
Zero background	$(100\ C_A)^2$	(80,120)	(60,140)	(30,70)
100	1.02 $(100\ C_A)^2$	(80,120)	(60,140)	(30,70)
50	1.04 $(100\ C_A)^2$	(80,120)	(59,141)	(30,70)
10	1.22 $(100\ C_A)^2$	(78,122)	(56,144)	(28,72)
5	1.48 $(100\ C_A)^2$	(76,124)	(51,149)	(26,74)
1	5.00 $(100\ C_A)^2$	(55,145)	(10,190)	(5,95)
0.5	13.0 $(100\ C_A)^2$	(28,170)	(-44,240)	(-22,122)
0.1	221 $(100\ C_A)^2$	(-200,400)	(-500,700)	(-247,347)
0.05	841 $(100\ C_A)^2$	(-480,680)	(-1100,1300)	(-530,630)
0.01	20,200 $(100\ C_A)^2$	(-2700,2900)	(-5600,5800)	(-1400,1500)
0.005	80,400 $(100\ C_A)^2$	(-5600,5800)	(-11,200,11,000)	(-5600,5700)

*95% tolerance interval for percent recoveries with assumed values for p and C_A [Tolerance limits = $100\,p \pm 1.96\sqrt{\mathrm{Var}\ (R_2)}$]

As can be seen from Table B-1, when $T/B = 1$, Var (R_2) is five times the zero-background value; when $T/B = 0.1$ Var (R_2) is about 221 times the zero-background value.

Suppose the background is not constant, i.e., B varies from sample to sample with

$$E(B) = \beta$$
$$V(B) = (\beta C_B)^2 \tag{B.6}$$

($100C_B$ is the between-sample coefficient of variation.) Then the variance of R_2 across samples is

$$V(R_2) = (100pC_A)^2[1+(K+1)^2+2C_B^2]K^{-2} \tag{B.7}$$

where $K=T/\beta$ (the ratio of spike level to average background concentration). It can be seen that $V(R_2)$ is larger than the true variance, $100pC_A$. Figure B-1 shows

$$SD(R_2)/SD(R_1) = [1+(K+1)^2+2C_B^2]^{1/2}/K \tag{B.8}$$

versus K for different between-sample coefficients of variation ($CV_B=100C_B$). The following points should be noted from the figure:

- For any C_B value $SD(R_2)$ is large when $K < 1$.
- C_B has little effect on $SD(R_2)$ when $K > 5$.
- Increasing C_B makes $SD(R_2)$ much larger when $K < 1$.

Between-sample variation generally is large for environmental samples (CV_B= 200% is not an unreasonable value). This variation cannot be controlled since measuring concentrations of unspiked samples (X values) is the laboratory's primary function.

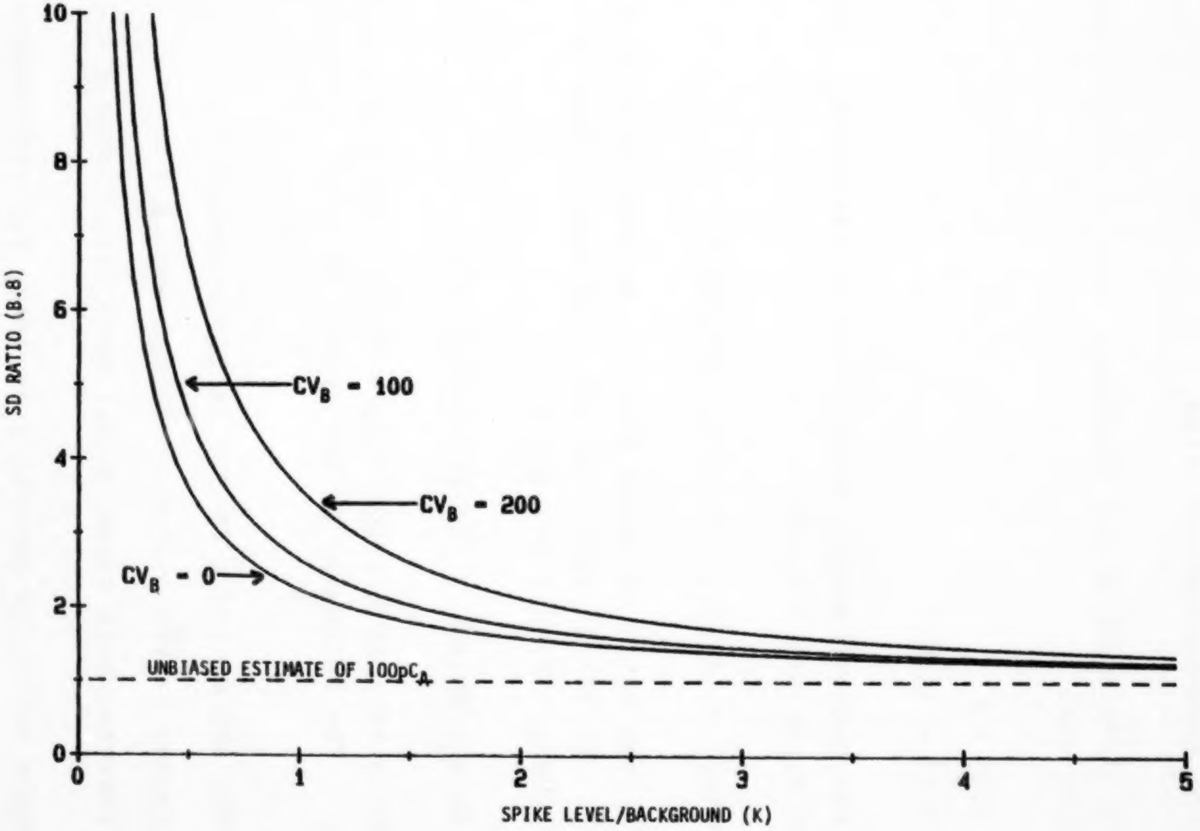


Figure B-1. Ratio of Actual Standard Deviation of Percent Recovery to Assumed Value - Fixed Spike Level

An example of the difficulty one can have with fixed spike levels appeared in a report on the effectiveness of wastewater treatment systems of 5 organic chemicals plants (2). A fixed spike level of 10,000 ppb was used in one laboratory. The measured background concentration on one sample was so high that the estimated K-value was 0.01; the calculated percent recovery for this sample was -7000 percent. This result obviously should have been excluded from percent recovery statistics but was not. It reflected poor spiking practice, not poor analytical performance.

Another example of potential misinterpretations of data from a spiking study can be found in reference (3). In this article, spiking studies were used to assess the performance of laboratories. The authors concluded that overall performance by the five labs in the study was poor. In one test, an unknown freshwater sample was analyzed with and without spikes of various minerals. The estimated spike/background ratios for the six minerals were as follows: 0.14, 1.1, 0.21, 5.0, 0.71, and 5.0. Some of the variability in recoveries attributed in the article to poor lab performance may have been due to the statistical properties of recoveries with low spike/background ratios.

Because of this potential problem the use of fixed-level spiking procedures is not recommended. Though they give an unbiased estimator of percent recovery, they can give results that are too variable to be useful for QC purposes.

NONZERO BACKGROUND - DEFINITION 3

In this definition,

$$R_3 = 100\ (Y-X)/hX$$

the spike level is a multiple (h) of the measured background concentration. The obvious motivation for this procedure is that it controls (within limits of the analytical method) the spike/background ratio.

Suppose that the true concentrations of original and spiked samples are B and hX+B and that analytical error is lognormally distributed. It has been shown (Reference (2)) that

$$E(R_3) = 100p\ (1+C_A{}^2/hp)$$

and (B.9)

$$V(R_3) = (100pC_A)^2[(hp+C_A{}^2+1)^2+(C_A{}^2+1)^3]/(hp^2)$$

Note that B is not in the equation for $V(R_3)$. Thus this percent recovery definition has an advantage over R_2 in that $V(R_3)$ does not depend on between-sample variation.

The ratio of standard deviations of R_3 and R_1 is

$$SD(R_3)/SD(R_1) = [(hp+C_A{}^2+1)^2+(C_A{}^2+1)^3]^{1/2}/hp \qquad (B.10)$$

This ratio is plotted in Figure B-2 as a function of hp and C_A. At hp=1 and C_A=0.25, $SD(R_3)$ is 2.34 times the actual value.

The following properties can be noted from (B.9) and Figure B-2:

- R_3 is a biased estimator of percent recovery whose bias increases with increasing C_A and decreasing h.
- $SD(R_3)$ increases as hp decreases or C_A increases.
- C_A has little effect on $SD(R_3)$ for hp > 5.

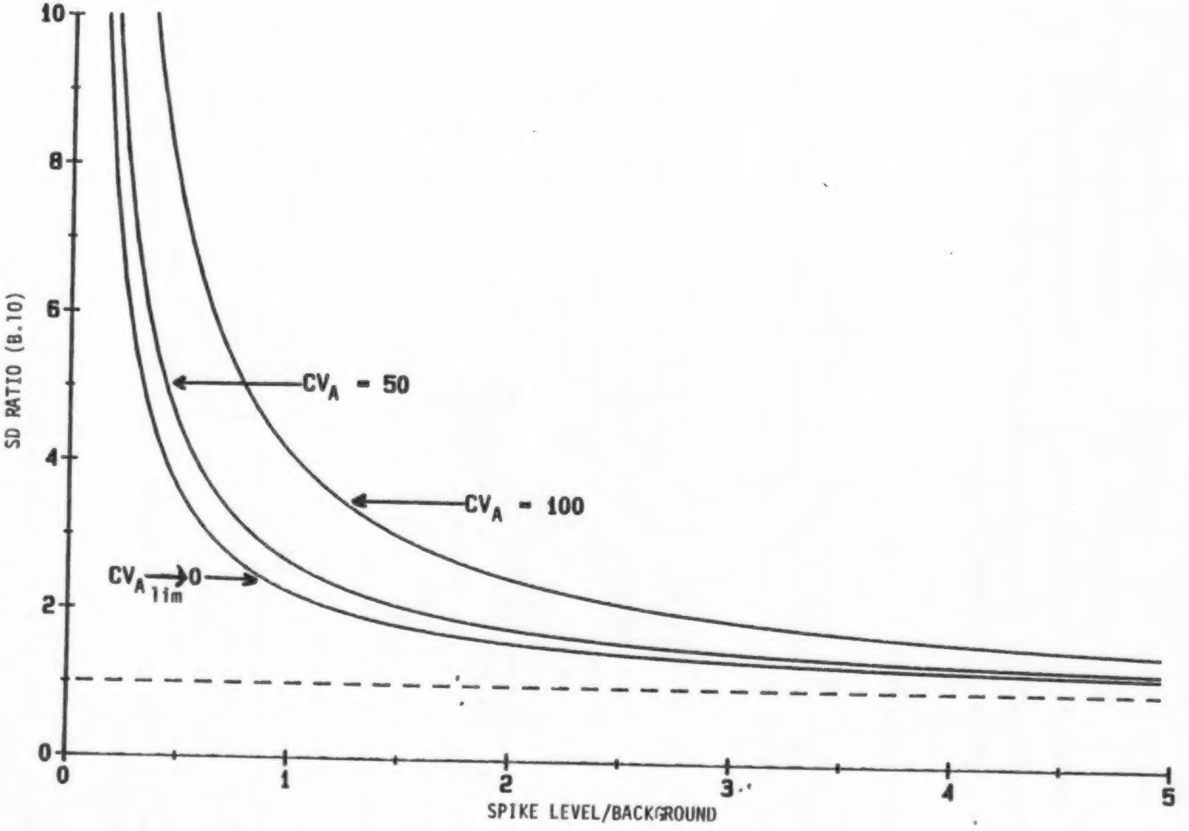


Figure B-2. Ratio of Actual Standard Deviation of Percent Recovery to Assumed Value - Proportional Spike Level

Note that $hp = E(Y-X)/E(X)$; that is, hp is the ratio of average spike level to average background concentration.

Based on the results discussed above, definition R_3 generally is preferrable to definition R_2. However, one must choose the h value properly or R_3 has deficiencies too. Note that p must be considered in choosing h because $V(R_3)$ is determined by hp, not by h alone. For example, h must be twice as large when $p = 1/2$ as when $p = 1$ to give the same $V(R_3)$.

COMMENTS AND CONCLUSIONS

1. It is desirable for a test method to have $P = 1$ (i.e., expected recovery of 100%) and Var (R) small, and that these parameters have the same value for all concentrations of interest.

2. In statistical terms, percent recovery data from zero background samples is easiest to interpret. The mean and variance of this data reflect the mean and variance of the measurement process. Such data should not be used naively to set control limits for percent recovery data from positive background samples, however, because of the statistical effects of positive background that were demonstrated in R_2 and R_2. (Matrix effects on method performance are another consideration.) Since contamination changes statistical properties of R_1, blank samples should always be analyzed along with spiked samples to check whether background concentration is truly zero.

3. The use of fixed ("blind") spike levels generally should be avoided. Though it is more convenient for the laboratory, this practice can result in low average spike/background ratios that drastically reduce the power of QC tests or cause false out of control signals (depending on how control limits are set). The fact that $V(R_2)$ depends on between-sample variation also reduces the usefulness of data from this procedure.

4. Proportional spiking is preferrable in the positive background case. But low h values must be avoided or R_3 will be biased and will be too variable to be useful for QC purposes. The EPA handbook (4) recommends $h=1$; Barnett and Youden (5) recommended $h=0.2$, 0.5 and 1. Larger values should be considered.

5. One way to choose spike levels is to cover the range of concentrations of interest, then find samples with background levels that are small compared to the chosen spike levels. If this is done, the recovery definition used will not affect conclusions. In some situations samples with low background levels may be difficult to obtain, however.

6. Another way to select spike levels is to make them multiples of estimated background levels, but small multiples can give misleading results. Samples still must be chosen carefully in this approach, or large multiples may lead to spike levels outside the range of interest.

7. Some analytical methods specify that one spike level should equal the initial background concentration. At this low multiple (k=1), R is more variable than at higher levels and s^2 overestimates the recovery variance. These properties could lead one to conclude that the recovery mean or variance of a method were different at low concentrations. This would needlessly complicate quality control procedures, since the apparent differences would be due to the spike and background levels used.

8. In some situations (e.g., studying the properties of a method near the detection limit), it may be difficult to obtain low background levels in the sample matrix of interest. Dilution of sample matrices should be considered. If it becomes necessary to perform spiking studies with a low spike/background ratio, the statistical properties of the recoveries should be considered in interpreting the results and in comparing them to results at other concentrations or in other matrices.

REFERENCES

1. Elder, R. S. and Provost, L. P., "Statistical Issues in QC for Trace Analysis," 28th Annual Fall Technical Conference: ASQC/ASA, London, Ontario, 1984.

2. Chemical Manufacturer's Association. CMA/EPA 5-Plant Study. Prepared by Engineering Science, Inc., Austin (1982).

3. Edwards, R. R., et al, "A Performance Evaluation of Certified Water Analysis Laboratories," Journal of Water Pollution Control Federation, 49, 1977, pp. 1704.

4. Environmental Protection Agency. Handbook for Analytical Quality Control in Water and Wastewater Laboratories. EPA-600/4-79-019, Office of Research and Development, Cincinnati (1979).

5. Barnett, R. N. and W. J. Youden. "A Revised Scheme for the Comparison of Quantitative Methods." American Journal of Clinical Pathology, 54 (190), pp. 454-462.